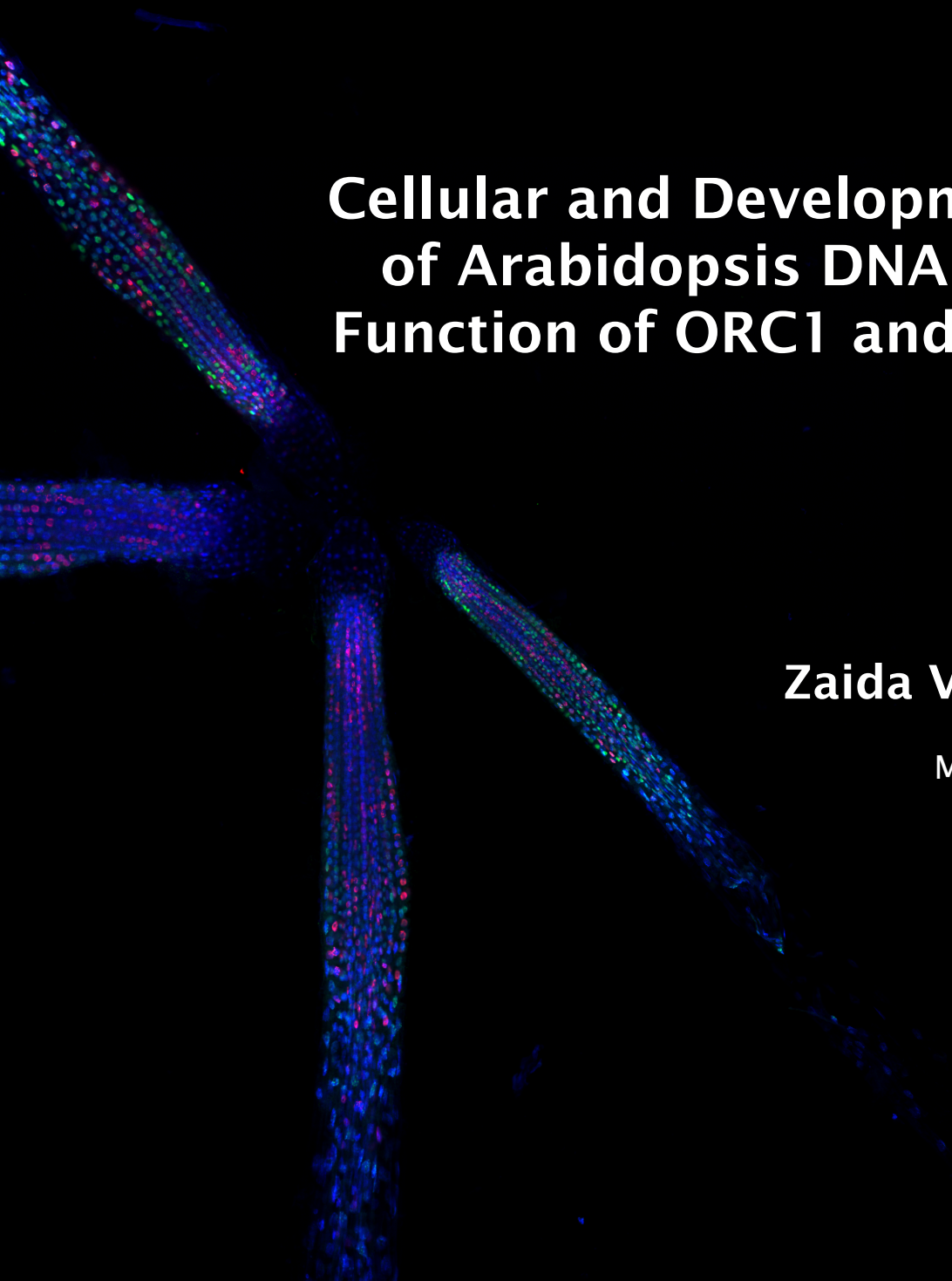
UNIVERSIDAD AUTONOMA DE MADRID

Facultad de Ciencias

Departamento de Biología Molecular



Cellular and Developmental Control of Arabidopsis DNA Replication: Function of ORC1 and Transposons

Zaida Vergara Pardilo

Madrid, 2017

Facultad de Ciencias
Departamento de Biología Molecular

Cellular and Developmental Control of Arabidopsis DNA Replication: Function of ORC1 and Transposons

Zaida Vergara Pardillo

Directed by:
Crisanto Gutierrez Armenta
Joana Sequeira Mendes

Madrid, 2017

El Dr. Crisanto Gutierrez Armenta, Profesor de Investigación del CSIC en el Centro de Biología Molecular Severo Ochoa

CERTIFICA

Que Doña Zaida Vergara Pardillo ha llevado a cabo bajo mi dirección los trabajos que conforman esta Tesis Doctoral con el título de "Cellular and developmental control of Arabidopsis DNA replication: function of ORC1 and transposons" en el centro de Biología Molecular Severo Ochoa.

El trabajo realizado reúne todas las condiciones requeridas por la legislación vigente, así como la originalidad y calidad científica para poder ser presentada y defendida con el fin de optar al grado de Doctor en Ciencias.

Este trabajo cuenta con el visto bueno del Dr. Luis Blanco Dávila, tutor académico en el Departamento de Biología Molecular de la Universidad Autónoma de Madrid.

Y para que conste donde proceda, firmo el presente certificado.

Madrid a 7 de Junio de 2017.

Dr. Crisanto Gutierrez Armenta

Director de la Tesis

Joana Sequeira Mendes, Dra. En Bioquímica,

CERTIFICA

Que Doña Zaida Vergara Pardillo ha llevado a cabo bajo mi co-dirección los trabajos que conforman esta Tesis Doctoral con el título de "Cellular and developmental control of Arabidopsis DNA replication: function of ORC1 and transposons" en el centro de Biología Molecular Severo Ochoa.

El trabajo realizado reúne todas las condiciones requeridas por la legislación vigente, así como la originalidad y calidad científica para poder ser presentada y defendida con el fin de optar al grado de Doctor en Ciencias.

Este trabajo cuenta con el visto bueno del Dr. Luis Blanco Dávila, tutor académico en el Departamento de Biología Molecular de la Universidad Autónoma de Madrid.

Y para que conste donde proceda, firmo el presente certificado.

Madrid a 7 de Junio de 2017.

Dra. Joana Sequeira Mendes

Co-directora de la Tesis

El Dr. Luis Blanco Dávila, Profesor de Investigación del CSIC en el Centro de Biología Molecular Severo Ochoa

CERTIFICA

Que Doña Zaida Vergara Pardillo ha llevado a cabo bajo la dirección del Dr. Crisanto Gutierrez Armenta los trabajos que conforman esta Tesis Doctoral con el título de "Cellular and developmental control of Arabidopsis DNA replication: function of ORC1 and transposons" en el centro de Biología Molecular Severo Ochoa.

El trabajo realizado reúne todas las condiciones requeridas por la legislación vigente, así como la originalidad y calidad científica para poder ser presentada y defendida con el fin de optar al grado de Doctor en Ciencias.

Este trabajo cuenta con mi visto bueno como tutor académico en el Departamento de Biología Molecular de la Universidad Autónoma de Madrid.

Y para que conste donde proceda, firmo el presente certificado.

Madrid a 7 de Junio de 2017.

Dr. Luis Blanco Dávila

Tutor Académico

This work was funded by Ministerio de Economía y Competitividad BFU2012-34821, BIO2013-50098-EXP, BFU2015-68396-R/FEDER. Zaida Vergara was a recipient of a predoctoral FPI (MINECO BES-2010-037158) fellowship.

A Lola, Félix y Arturo,

“The nitrogen in our DNA, the calcium in our teeth, the iron in our blood, the carbon in our apple pies were made in the interiors of collapsing stars. We are made of starstuff.”

Carl Sagan, Cosomos

“Nothing in Biology makes sense except in the light of evolution”

Theodosius Dobzhansky

Agradecimientos

Muchas gracias a todos los que habéis hecho esta aventura posible.

Primero, gracias Crisanto, por brindarme la oportunidad de trabajar en tu laboratorio todos estos años y ayudar a convertirme en la científica que ya me creo que soy. Gracias por las horas de discusión científica, por no decir no a las probatinas que se me ocurrían y por el gran apoyo durante la escritura. Ha sido un largo camino, pero como dicen todo llega y afortunadamente estamos empezando a ver los frutos del trabajo y el esfuerzo, ¡qué más da eso del 13!

Obrigada Joana, porque (cronológicamente) fuiste amiga antes que "jefa". Mil gracias por hacer *pull down* conmigo y más de mil gracias por los nachos con guacamole acompañados de una rica margarita, aún nos quedan más de esos seguro. No sólo me has enseñado a llevar los retos dentro del lab, también fuera, y eso no estaba en el trato, jajaja, tu me entiendes. Gracias por enseñarme a ser crítica sin ser escéptica con mis resultados, que el trabajo duro tiene recompensa y que a veces es mejor dar un paso atrás y pensar las cosas con calma y una cervecita.

Muchos habéis pasado por el lab 308 en estos años. Gracias Celina por guiarme en los primeros pasos, Maribel por enseñarme Gateway, Marta por la compañía en la última U, Nuria por poner radio 3 y las risas de última hora del día, a Irene, las dos Elenas, Adriana, Alex, María D. y Sergio por su alegría y experiencia, thanks Martina for the BSA-tips :D, David por las risas y la tranquilidad que transmites, Ainhoa por todos los post-its llenos de cosas bonitas y todos los cafés y ánimos incluso a día de hoy, Aitor por el cartel de las flores de estufa, jejejej, Clara por el torbellino de energía que transmites, María FM por todos los truquitos con las plantas, Carla gracias por los brownies, aprenderé a hacerlos aún a riesgo de comerme uno al día, Béné por las charlas de política, ojalá el país se arregle pronto, Victoria por las historias de viajes, los mapas y las recomendaciones, y Sofia, obrigada amiga, porque no sólo me has recordado que tenía que respirar antes de entrar en pánico con la escritura y las fechas, además tienes paciencia para enseñarme portugués y hacer chistes de kinasas ;D Gracias Iluminada por la sonrisa de cada tarde y por adoptarme, estos años no habrían sido iguales si hubiese estado en otra planta. Gracias a don José Cuervo por el apoyo en los puntos más altos y bajos durante estos años.

Gracias a María Gómez y a su lab por toda la discusión científica. Ricardo, ¡qué ya acabamos! Pero sólo la tesis, que amistad aún queda para rato. Josemi, Alba, Gonzalo, Cristina y Laura gracias por las risas, los cafés y las cervecitas. Rodri, la experiencia no habría sido la misma sin ti, para empezar probablemente no habría acabado este libro, así que elige la parte que más te guste que es tuya. Gracias por venir a Montpellier y enseñarme Edimburgo, los cafés, las risas y los consejos.

A los vecinos del lab de Encarna, Jorge, Alfonso, Ted, David, Noemí, Javi, Gloria, Rosa, Charo y Azmane, gracias por las risas en las comidas y la ayuda ocasional. Gracias Ramón por la paciencia con la bioinformática y a los servicios del CBM, en especial a Ángeles, Carmen, Maite y Vero.

Merci beaucoup á Etienne Schowb. I had a great time in your lab. Thanks a lot to Vincent for teaching me combing and to the rest of the lab Marjorie, Elisabeth and Léon. Gracias a Andrés por cuidarme como a una hermana y presentarme a toda la gente de Montpellier, sobre todo a Brenda, mi hermana mexicana, ¡nos vemos pronto!. Gracias a vosotros tres meses fueron más como tres años de experiencias.

A mis biólogos favoritos, Ana, Diego, Patri, Lucía, Jeni, Chus, Irene, Nerea, Bea, Lara, Alberto, David, Ester, Inés, José, Vivi, gracias por encontrar huecos para seguir quedando y contarnos las cosas, me encanta que aunque pase el tiempo la confianza y las risas perduren y sigamos planeando viajes.

Gracias Héctor, Araceli y Azahara, por todo el apoyo estos años y porque siempre estáis dispuestos a un skype, una cena o una llamada cuando hace falta. Gracias Clara por enseñarme lo bonito de correr, espero que donde estés seas feliz.

Muchas gracias a toooooda mi familia que me habéis escuchado hablar de raíces, de ORC, de cromatina… ¡y con atención! Sandra, Aitor y Sergio, gracias por venir a Montpellier, y hacerme la compra, jejeje. Fue un finde bien divertido y no el único. Gracias padres por escuchar mis posters y ayudarme a medir raíces. Arturo, gracias por leerte la tesis eso sí que tiene mérito, no puedo decir más.

Y una es la familia que te toca y otra es la familia que eliges, gracias vicalvareños por muchos años de risas, emociones y experiencias. Almu, Ana Belén, Arantxa, Cristina, Dani, Diana,

Estherci, Felipe, Javi, José, Maribel, Noelia, Vike, Tamara, Pablo y Beita, en Madrid o en cualquier punto del planeta.

Index

A

AAA+: ATPases Associated with various cellular activities

ABRC: Arabidopsis Biological Resource Center

ABRE-like: ABA-RESPONSIVE ELEMENTS-LIKE

ACS: ARS consensus sequence

ACT2: ACTIN2

ANOVA: Analysis of Variance

AP-2: APETALA-2

APC: ANAPHASE PROMOTING COMPLEX

ARF: AUXIN RESPONSE FACTOR

ARS: Autonomous replicating sequence

ATXR: ARABIDOPSIS TRITHORAX-RELATED

B

BAC: Bacterial Artificial Chromosome

BAH: Bromo-Adjacent Homology

BCA: Bicinchoninic acid

bHLH: basic Helix Loop Helix

BRCA1: BREAST CANCER SUSCEPTIBILITY1

BrdU: 5-bromo-2'-deoxyuridine

BSA: Bovine Serum Albumin

C

CAF-1: CHROMATIN ASSEMBLY FACTOR 1

CAK: CYCLIN DEPENDENT KINASE (CDK) ACTIVATING KINASE

CCS52: CELL CYCLE SWITCH52

CDC45: CELL DIVISION CYCLE45

CDC6: CELL DIVISION CYCLE6

CDK: CYCLIN DEPENDENT KINASE

CDT1: CDC10 DEPENDENT TRANSCRIPTION1

CEI: Cortex-Endodermis Initials

CENH3: CENTROMERIC H3

CENP-A: Centromere protein A

ChIP: Chromatin Immunoprecipitation

CKI: CYCLIN DEPENDENT KINASE (CDK) INHIBITOR

CLE40: CLAVATA3/ESR-RELATED40

CMG: CDC45-MCM-GINS

CMT: CROMOMETHYLASE

Col-0: Columbia

CP190: Centrosome-associated zinc finger protein 190

CRISPR-Cas9: Clustered Regulatory Interspaced Short Palindromic Repeats-Cas9

CS: Chromatin State

CTAB: cetyl trimethylammonium bromide

CTCF: CCCT-binding factor

CYC: CYCLIN

D

DAPI: 4',6-diamidino-2-phenylindole

DDK: Dbf4-dependent kinase

DDM1: DECREASED DNA METHYLATION1

DMSO: dimethyl sulfoxide

DP: DIMERIZATION PARTNER

dps: days post sowing

DRM2: DOMAINS REARRNGED METHYLASE2

E

E2F: ADENOVIRUS E2 PROMOTER BINDING FACTOR

EDTA: Ethylenediaminetetraacetic acid

EdU: 5–ethynyl–2'–deoxyuridine

EGTA: Ethylene glycol-bi(β-aminoethyl ether)-N,N,N',N'-tetraacetic acid

ELI: Epidermis-Lateral root cap Initials

ERF115: ETHYLENE RESPONSE FACTOR115

F

FAS1: FASCIATA1

FBL17: F BOX-LIKE17

FM4-64:(*N*-(3-Triethylammoniumpropyl)-4-(6-(4-(Diethylamino)Phenyl)Hexatrienyl) Pyridinium Dibromide)

G

G3GFP: G3 GREEN FLUORESCENT PROTEIN

G4: G-quadruplexes

GAPC2: GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C2

GATA: GATA TRANSCRIPTION FACTOR

GEM: GLABRA2 (GL2) EXPRESSION MODULATOR

GFP: GREEN FLUORESCENT PROTEIN

GINS: From the Japanese 5-1-2-3 (*go-ichi-ni-san*) Sld5-Psf1-Psf2-Psf3 complex

GRAS: GIBBERELLIC-ACID INSENSITIVE (GAI), REPRESSOR OF GAI (RGA) and SCARECROW (SCR)

GUS: β-Glucuronidase

H

HBO1: Histone acetyltransferase Binding to ORC

HOMER: Hypergeometric Optimization of Motif EnRichment

HP1: HETEROCHROMATIN PROTEIN1

HRP: HORSERADISH PEROXIDASE

HS: Horse Serum

HU: Hydroxyurea

I

ICC: Immunocytochemical assay

IHC: Immunohistochemical assay

K

KRP: KINASE INTERACTING PROTEIN (KIP)-RELATED PROTEIN

KYP: KRYPTONITE

L

LB: Luria-Bertani

LINE: Long interspersed elements

LRC: Lateral Root Cap

LSM: Laser Scanning Microscope

LTR: Long terminal repeats

M

MACS: Model-based Analysis of ChIP-Seq

MCM: MINICHROMOSOME MAINTENANCE

MES: 2-(N-morpholino)ethanesulfonic acid

MET1: METHYLTRANSFERASE1

MG132: Z-Leu-Leu-Leu-aldehyde

MGS: Meier-Gorlin Syndrome

MLN4924: Pevonedistat

MOPS: 3-(N-morpholino)propanesulfonic acid

mRFP: monomeric RED FLUORESCENT PROTEIN

MS: Murashige and Skoog

MTSB: Microtubules stabilizing buffer

MYB36: MYB DOMAIN PROTEIN36

N

NAP1: NUCLEOSOME ASSEMBLY PROTEIN1

NEDD8: NEURAL PRECURSOR CELL EXPRESSED, DEVELOPMENTALLY DOWN-REGULATED8

NGS: Next Generation Sequencing

O

ORC: ORIGIN RECOGNITION COMPLEX

ORI: Origin of DNA replication

P

PARP1: POLY (ADP-RIBOSE) POLYMERASE1

PBS: Phosphate-Buffered Saline

PBST: PBS-Tween-20

PCNA: PROLIFERATION CELL NUCLEAR ANTIGEN

PCR: Polymerase Chain Reaction

PHD: Plant Homeodomain

PIN: PIN-FORMED

PIPES: Piperazine-N,N'-bi(2-ethanesulfonic acid)

PLT: PLETHORA

PMSF: phenylmethanesulfonylfluoride

PNK: T4 Polynucleotide Kinase

PRC: POLYCOMB REPRESIVE COMPLEX

pre-IC: pre-Initiation Complex

pre-RC: pre-replication complex

Q

QC: Quiescent Center

qPCR: quantitative PCR

QR: Quick Response code

R

RAD51: RADIATION SENSITIVE51

RAM: Root Apical Meristem

RAV1-A: RELATED TO ABI3/VP1

RBR: RETINOBLASTOMA RELATED

RdDM: RNA-dependent DNA methylation

ROI: Region of Interest

RPA: REPLICATION PROTEIN A

RT: Reverse Transcriptase

S

SAM: Shoot Apical Meristem

SCF: S PHASE KINASE-ASSOCIATED PROTEIN 1 (SKP1)/CULLIN/F-BOX

SCR: SCARECROW

SDS: Sodium Dodecyl Sulfate

SHR: SHORTROOT

SHY2: SHORT HYPOCOTYL2

SICER: Spatial Clustering for Identification of ChIP-Enriched Regions

SIM: SIAMESE

siRNA: short interfering RNA

SKP2: S PHASE KINASE-ASSOCIATED PROTEIN 2

SNS: Short Nascent Strands

SORLIP: SEQUENCE OVER-REPRESENTED IN LIGHT-INDUCED PROMOTERS

SUVH: SU(VAR)HOMOLOG

T

TAD: Topologically associating domain

T-DNA: Transgenic DNA

TE: Transposable element

TF: Transcription Factor

TF: Transcription Factor

T-PIC: Tree shape Peak Identification for ChIP-seq

TREX-2: TRANSCRIPTION-COUPLED EXPORT-2

TSS: Transcription Start Site

U

UPB1: UPBEAT1

UTR: Untraslated Region

W

WB: Western blot

WEE1: WEE1 KINASE HOMOLOG

WH: Winged Helix domain

WOX5: WUSCHEL-RELATED HOMEOBOX5

X

X-Gluc: 5-bromo-4-chloro-3-indoyl-β-D-glucuronide

Abstract

Abstract

Plant stem cells are organized into meristems that actively divide throughout the lifespan of the plant to produce new organs. Once cells stop proliferating they frequently enter the endocycle program, which in plants is usually linked to the differentiation process. Reliable genome duplication helps maintaining genomic stability of meristematic and endocycling cells. This process starts at discrete sites called origins of DNA replication (ORIs), marked by a set of proteins that form the pre-replication complex. We focused our analysis on the largest subunit of the origin recognition complex (ORC1), which plays a key role in ORI licensing. In addition, ORC1 is tightly controlled to avoid re-replication problems in a species-specific manner.

We have combined confocal microscopy techniques with analysis of translational fusion constructs and mutant plants to determine that the function, dynamics and regulation of the two Arabidopsis ORC1 proteins are different during root organogenesis. While ORC1a plays a role in the maintenance of heterochromatin in endocycling cells, ORC1b functions in ORI licensing during proliferation. Our analysis shows that ORC1b associates with euchromatin and heterochromatin in proliferating cells. The protein bounds to chromatin since G2 until the G1/S transition. Upon S-phase entry the E3-ubiquitin ligase SCF^{FBL17} recognizes ORC1b triggering its degradation. Plants lacking ORC1b are hypersensitive to aphidicolin, most likely due to having less licensed ORIs. On the contrary, ORC1a is only present in endocycling cells and preferentially associates with heterochromatin. Although ORC1a is only present at the G-phase of the endocycle, its degradation is independent of the ubiquitin proteasome pathway. The mark H3K27me1 is severely diminished at chromocenters of endocycling cells of plants lacking ORC1a pointing to a role of the protein in the establishment of the heterochromatin in this group of cells.

In addition, we focused on defining the features associated with ORIs in heterochromatin. Arabidopsis ORIs preferentially colocalize with genes, but in pericentromeric gene-poor domains, a large proportion associate with transposable elements (TEs). ORI-TEs colocalize almost exclusively with retrotransposons, in particular of the Gypsy family. Opposite to ORI-genes, ORI-TE activity occurs independently of TE expression. In addition, ORI-TEs maintain high levels of the repressed heterochromatin marks H3K9me2 and H3K27me1. We have found a specific chromatin signature of ORI-TEs, defined by GC-rich heterochromatin. Importantly, TEs with active ORIs contain a local GC content higher than the TEs lacking them. Our results lead us to conclude that ORI colocalization with retrotransposons is determined by a specific chromatin landscape and their transposition mechanism based on transcription.

Resumen

Resumen

Las células madre vegetales se organizan en meristemos, dónde se dividen activamente a lo largo de la vida de la planta produciendo nuevos órganos. Una vez que las células dejan de proliferar, entran en el programa de endociclo, ligado a la diferenciación celular. La correcta duplicación del genoma es esencial para mantener la estabilidad genómica de las células proliferantes y en endociclo. La replicación del ADN comienza en sitios discretos llamados orígenes de replicación del ADN (ORIs), marcados por un conjunto de proteínas que forman el complejo pre-replicativo. Nuestro estudio se ha centrado en la subunidad mayor del complejo de reconocimiento del origen (ORC1), dado que juega un papel importante en el licenciamiento de los ORIs. Además, ORC1 está estrictamente controlada para evitar problemas de re-replicación, si bien la regulación es específica para cada organismo.

Hemos combinado técnicas de microscopía confocal y análisis de plantas mutantes o con fusiones traduccionales en los dos genes *ORC1* de Arabidopsis y hemos determinado que su función, dinámica y regulación son diferentes durante el desarrollo de la raíz. Mientras que ORC1a juega un papel en el mantenimiento de la heterocromatina en las células en endociclo, ORC1b funciona en el licenciamiento de los ORIs durante la proliferación. ORC1b se asocia con eucromatina y heterocromatina en células en proliferación. La proteína se une a la cromatina en G2 y permanece unida hasta la transición G1/S, cuando la E3 ubiquitina ligasa SCF^{FBL17} reconoce ORC1b desencadenando su degradación. Las plantas que carecen de ORC1b son hipersensibles a afidicolina, debido a que tienen menos ORIs licenciados. Por su parte, ORC1a sólo está presente en células en endociclo y preferiblemente asociada con heterocromatina. Aunque ORC1a sólo está presente durante la fase G del endociclo, su degradación es independiente del proteasoma. La falta de ORC1a produce una severa disminución de la marca H3K27me1 en los cromocentros de células en endociclo, indicando que la proteína juega un papel en el establecimiento de la heterocromatina en este grupo de células.

Además, hemos definido las características asociadas a los ORIs de la heterocromatina. Los ORIs en Arabidopsis se localizan principalmente en genes, pero en la zona pericentromérica pobre en genes, una gran parte de los ORIs se asocian a elementos transponibles (TEs). ORI-TEs colocalizan principalmente con retrotransposones, en concreto de la familia Gypsy. Al contrario de lo que sucede con los ORIs en genes, la actividad ORI-TE es independiente de la expresión del TE y sucede en presencia de altos niveles de las marcas de heterocromatina H3K9me2 y H3K27me1. Hemos encontrado que la característica que define a los ORI-TEs es la heterocromatina rica en GC. Así, los TEs con ORIs activos poseen un contenido GC local mayor que los TEs que carecen de ellos. Nuestros resultados llevan a la conclusión de que la colocalización de los ORIs con los retrotransposones se determina por su mecanismo de transposición basado en la transcripción y un paisaje de cromatina específico.

1. Introduction

1.1 Plants as a model to study cell cycle and development

Plant and animal lineages separated during evolution 1,6 billions years ago (Parfrey et al., 2011). Although they currently represent different living strategies, exhibiting unique characteristics at the developmental and organogenesis level, basic molecular and cellular processes are frequently conserved, for instance, protein synthesis or cell division cycle regulation (see below). Initial plant developmental studies date from the 18th century, when meristems were first described ((Wolff, 1774), reviewed in (Prunet and Meyerowitz, 2016)). However, it was only 50 years ago when the information retrieved using specific mutants was coupled with the developmental knowledge (Prunet and Meyerowitz, 2016). In all these studies cell proliferation was found to be crucial for developmental programs. Consequently, plant cell cycle studies that started focusing on the control of cell cycle progression lead to the current understanding of cell proliferation control in the context of organogenesis (reviewed in (Gutierrez, 2016)). From that point, plants have been used extensively as a model to study organogenesis, which in plants is post-embryonic. Plants are able to produce new organs throughout their lifespan, such as leaves, flowers or roots, in response to environmental cues. This gives them a huge phenotypic plasticity and offers the possibility of studying cell proliferation and development in the adult organism (Gutierrez, 2016; Morao et al., 2016; Xiao et al., 2017). In general, plant hormones, cell cycle related genes, the proteasome pathway and transcription factors would control organogenesis processes in plants, establishing different regulatory layers that interact one with each other. Stem cells are undifferentiated cells able to divide and produce new tissues. While adult animals only have multipotent stem cells that function to maintain tissue homeostasis, adult plants contain pluripotent stem cells organized into meristems (Heidstra and Sabatini, 2014). During early plant embryogenesis, two meristems are specified, the shoot and the root meristems. After germination, the stem cell niches of both meristems are maintained and give rise to the above ground organs and the root system, respectively. The root system grows underground and provides structural support, water and nutrients to the plant. Remarkably, plants tolerate mutations in essential developmental genes that are lethal in other eukaryotes (Pikaard and Mittelsten Scheid, 2014), and manage to overcome higher levels of DNA damage (Manova and Gruszka, 2015), making this system even more interesting to study developmental processes.

Arabidopsis thaliana is a plant model organism very useful to study cell cycle and developmental processes. First, the genome is quite small (125 Mb organized in 5 nuclear chromosomes, plus the chloroplast and mitochondrial DNA), already sequenced (AGI, 2000) and very well annotated (Cheng et al., 2017; Lamesch et al., 2012). Second, it is a small plant

with a short life cycle (around two months) and produces a large number of seeds in every generation. Third, Arabidopsis is easy to transform and there are several mutant collections available, simplifying functional studies. In addition, the CRISPR-Cas9 technology has been successfully adapted to plants, increasing enormously the experimental possibilities (Puchta, 2017). Finally, Arabidopsis roots are highly organized at the cellular level and easy to image by confocal microscopy, being an extraordinary developmental model system that has been studied for the last 30 years (Petricka et al., 2012).

1.2 The Root Apical Meristem

The Arabidopsis root apical meristem (RAM; Fig. 1A) is specified during early embryogenesis when the AP2-domain *PLETHORA* (*PLT*) transcription factors (TFs) are expressed (Aida et al., 2004). The RAM stem cell niche (Fig. 1B) consists of different initial stem cells located around two organizing cells, a situation that resembles stem cell niches found in animals (Li and Clevers, 2010; Scheres, 2007). A few slowly dividing cells made up the quiescent center (QC) that prevent the differentiation of the surrounding cells (van den Berg et al., 1997). QC cells act as a reservoir of stem cells (or intact DNA) and only divide frequently upon damage of the nearby stem cells (Heyman et al., 2013), for example, after accumulation of the ETHYLENE RESPONSE FACTOR115 (ERF115) TF, which is negatively regulated through the ubiquitin proteasome pathway and positively by the plant hormone brassinosteroid (Heyman et al., 2013).

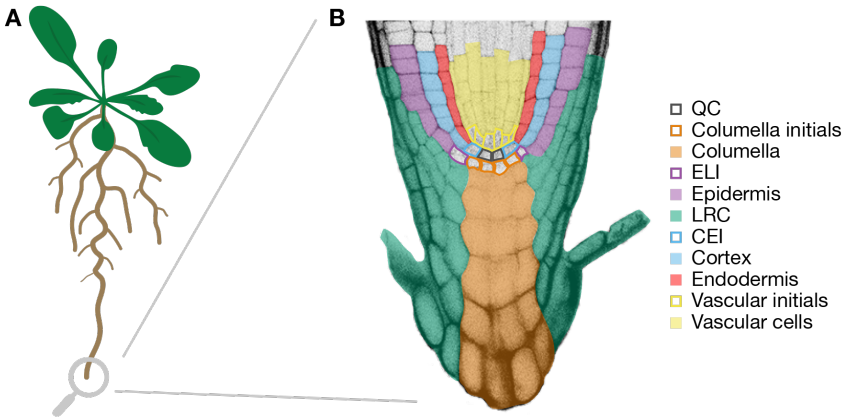


Figure 1 – Arabidopsis root stem cell niche. (A) In an adult Arabidopsis plant the root apical meristem is localized at the tip of the primary root. **(B)** The stem cells are organized surrounding the organizing cells of the quiescent center (QC). ELI: Epidermis-Lateral root cap Initials; LRC: Lateral Root Cap; CEI: Cortex-Endodermis Initials.

The initial stem cells derived from the QC produce different cell types (Fig. 1B): distal columella cells originate from the columella stem cells; epidermis and lateral root cap arise

from the epidermis-lateral root cap initial cells; ground tissue (cortex and endodermis) results from an anticlinal division of the cortex-endodermis initial cells (Heidstra and Sabatini, 2014; Petricka et al., 2012); phloem and xylem tissues derive from the vascular stem cells (De Rybel et al., 2016).

Besides specific TFs there are other layers of regulation responsible for the RAM stem cell niche identity, one of them being the hormonal regulation. The plant hormone auxin is accumulated at the QC thanks to the polar transport of this growth regulator through the *PIN* family of transmembrane proteins (Petrasek and Friml, 2009). In particular, *PIN1*, *PIN3*, *PIN4* and *PIN7* contribute to the acropetal movement of the auxin (from the shoot to the RAM) while *PIN2* transports auxin back to the shoot in a basipetal manner (Petrasek and Friml, 2009). Auxin accumulation at the QC triggers the expression of *PLT1* and *PLT2* (Aida et al., 2004). Remarkably, PLTs regulate *PIN* expression resulting in a positive loop that controls root patterning through preservation of high auxin and PLT levels at the stem cell niche (Blilou et al., 2005).

In parallel, the pathway dependent on the GRAS family TFs *SHORT-ROOT* (*SHR*; (Benfey et al., 1993)) and *SCARECROW* (*SCR*; (Sabatini et al., 2003)) maintains QC identity. SHR, which is expressed in the stele (vascular tissue), moves to the QC and activates SCR (Helariutta et al., 2000; Nakajima et al., 2001; Sabatini et al., 2003) forming a complex (Cui and Benfey, 2009). SHR-SCR cell-autonomously defines the organizer identity of the QC cells (Helariutta et al., 2000; Sabatini et al., 2003). It has been reported that the inhibition of QC divisions is mediated through an SCR-RETINOBLASTOMA (RBR) protein network (Cruz-Ramirez et al., 2013). In addition, SHR-SCR action controls the anticlinal division of the cortex-endodermis initial cells to generate cortex and endodermis ground tissues (Hirsch and Oldroyd, 2009). This process is dependent on the activation of CYCD6;1 by SHR-SCR, that would trigger the phosphorylation of RBR allowing the division (Sozzani et al., 2010).

The presence of another imbricated factor, *WUSCHEL-RELATED HOMEOBOX5* (*WOX5*; (Haecker et al., 2004; Sarkar et al., 2007)), preserves the undifferentiated state of the stem cells (Sarkar et al., 2007). The CLAVATA3/ESR-RELATED40 (CLE40) peptide is expressed in the differentiated columella cells and prevents WOX5 activity by promoting cell differentiation (Sarkar et al., 2007).

1.3 Developmental Zones of the Root

Root cells originated from the stem cells will divide a number of cycles, which is not well known and depends on the cell layer and the plant species (in Arabidopsis between 3 and 4

times), creating a pool of undifferentiated cells (Fig. 2A). Mitotic competence in the RAM decreases as cells distance from the QC (Ivanov and Dubrovsky, 2013). Two zones inside the RAM can be distinguished depending on their differential proliferation capacity: the proliferation zone, close to the stem cell niche where all cells proliferate and divide, and the transition zone, the shootward half where mitotic divisions are rare (Ivanov and Dubrovsky, 2013).

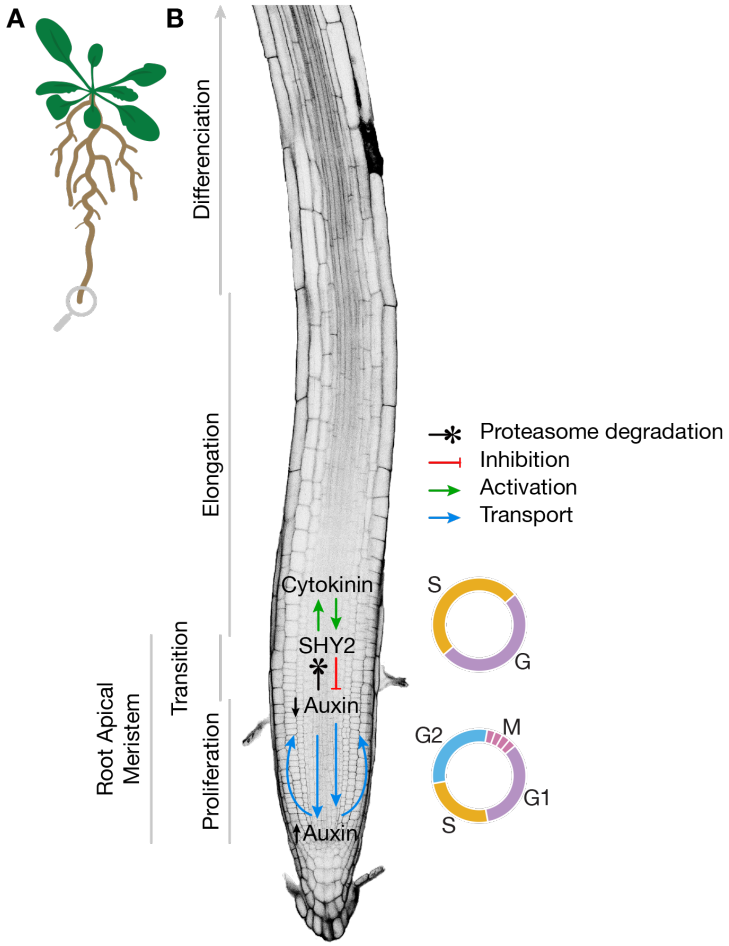


Figure 2 – Developmental zones in the Arabidopsis root. (A) In an adult Arabidopsis plant the root apical meristem is localize at the tip of the primary root. **(B)** The root apical meristem is divided in two different domains, the proliferation domain were cells actively divide, and the transition domain, where mitoses are less frequent and endocycle program starts. Outside the meristem, in the elongation zone, cells rapidly grow and once they reached their final size they start the differentiation process at the differentiation zone of the root. The control of the regulatory network between auxins and cytokinins is shown. Cytokinins promote the expression of SHY2 that inhibits the auxins, while auxins trigger the proteasome degradation of SHY2, which control cytokinins.

Once cells exit the RAM, they start a rapid cell expansion at the elongation zone. The visual boundary defined by the position where cells start to elongate determines the end of the meristem, which is not the same for each cell layer in the root. The plant hormones auxin and cytokinin determine the balance between proliferation and differentiation at the end of the RAM (Fig. 2B; (Schaller et al., 2015)). Cytokinin expression is enhanced at the transition to elongation zone where it induces the expression of SHORT HYPOCOTYL2 (SHY2) protein (Dello Ioio et al., 2008). SHY2 inhibits auxin response TFs (ARFs) reducing the levels of auxin at the end of the meristem and promoting the accumulation of auxins at the QC (Barrada et al., 2015; Dello Ioio et al., 2008). At the same time, auxins promote the proteasome degradation of SHY2

(Dharmasiri et al., 2003; Tian et al., 2002). SHY2 also promotes cytokinin biosynthesis, consequently, auxins limit cytokinins at the end of the meristem. Moreover, the boundary between RAM and the elongation zone has been related to the balance of reactive oxygen species, which is controlled by *UPBEAT1* (*UPB1*), a bHLH TF (Tsukagoshi et al., 2010) and *MYB36*, a MYB TF (Fernandez-Marcos et al., 2017; Liberman et al., 2015). The RAM size is not only dependent on hormones and TFs: chromatin associated changes are also important. For example, the transition from cell proliferation to differentiation coincides with the massive eviction of the canonical histone H3.1 and its replacement with the variant histone H3.3 (Otero et al., 2016). Once cells have reached their final length, they differentiate and acquire their specific functions, leading to a new zone in the root: the differentiation zone (Fig 2B).

The cell cycle is a highly conserved and regulated process with the purpose of giving rise to two daughter cells (Fig. 3). In general, it will produce two identical cells in order to increase the number within a population. However, in some cases, especially associated to morphogenesis and differentiation processes, an asymmetric division would lead to two daughter cells with different structural or functional properties (reviewed in (Desvoyes et al., 2014)), for instance, the anticlinal division of the cortex-endodermis initial cells that leads to the cortex and endodermis ground tissue (Hirsch and Oldroyd, 2009).

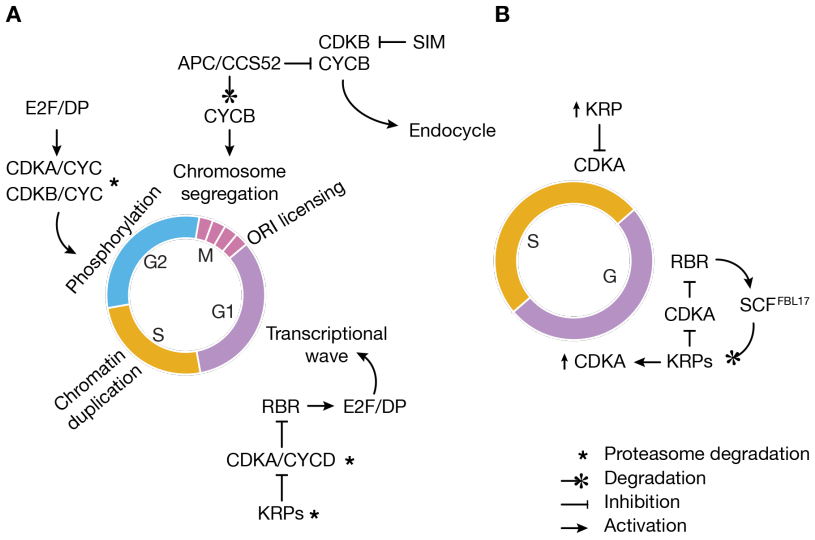


Figure 3 – Regulatory mechanism of Arabidopsis cell cycle and endocycle progression. Multiple Cyclins (CYC), Cyclin Dependent Kinases (CDK), activators and inhibitors, in coordination with the ubiquitin proteasome pathway regulate the progression. **(A)** During cell cycle S-phase CDK (S-CDK) and mitotic CDK (M-CDK) activity allow G1/S and G2/M transitions, respectively. **(B)** Endocycle onset is triggered by suppression of M-CDK activity. Endocycle progression is achieved by an oscillation between high S-CDK and high CKI (CDK Inhibitor) activities. KRP: KINASE INTERACTING PROTEIN (KIP)-RELATED PROTEIN; RBR: RETINOBALSTOMA RELATED; APC: ANAPHASE PROMOTING COMPLEX; CCS52: CELL CYCLE SWITCH52; SIM: SIAMESE; SCF^{FBL17}: SKP1/CULLIN/F-BOX FBL17.

The activity of various cyclins (CYC) and CYCLIN DEPENDENT KINASES (CDKs) allows cell cycle progression (Fig. 3A; (Gutierrez, 2009)). There are cell cycle activators, such as the CDK ACTIVATING KINASES (CAK), and specific inhibitors, such as the KINASE INTERACTING PROTEIN (KIP)-RELATED PROTEIN (KRPs) that will determine the active or inactive state of the CDK/CYC complexes. The availability of most of the cell cycle proteins is controlled through the ubiquitin proteasome pathway. In plants, this regulation layer is even more complex, due to the increased number of proteasome components, for instance, there are more than 600 F-box proteins (Gagne et al., 2002).

During the gap between mitosis and DNA replication (G1) cells growth and duplicate the organelles. In addition, the sites where DNA replication will start (ORIs) are marked by a set of proteins that form the pre-replication complexes (pre-RCs). This process is called licensing of the ORIs (see below). The G1 associated transcriptional wave is dependent on the RBR/E2F pathway (Berckmans and De Veylder, 2009; Gutierrez et al., 2002) and conserved between plants and animals (Weinberg, 1995). CDKA/CYCD complexes phosphorylate RBR releasing the TFs E2F/DP that trigger the expression of specific genes. Among those are included genes required for genome duplication in S-phase such as pre-RC factors *ORC* (*ORIGIN RECOGNITION COMPLEX*; (Diaz-Trivino et al., 2005)), *CDC6* (*CELL DIVISION CYCLE6*; (Castellano et al., 2001)) or *MCM3* (*MINICHROMOSOME MAINTENANCE3*; (Stevens et al., 2002)). Another regulatory layer is achieved by the inhibition of the CDK/CYC activity by the specific CDK inhibitors KRPs (De Veylder et al., 2001). In addition, CYCs, E2F/DP TFs and KRPs availability is controlled by the ubiquitin proteasome pathway (reviewed in (Gutierrez, 2009)).

In S-phase, the chromatin, the macromolecular complex formed by the association of the genomic DNA with histones and non-histone proteins, is duplicated. Accordingly, thousands of ORIs will fire coordinately to initiate the copy of the DNA and a set of histone chaperones will incorporate nucleosomes in a replication dependent manner, for instance the CHROMATIN ASSEMBLY FACTOR (CAF-1) that deposits H3.1-H4 dimers (Ramirez-Parra and Gutierrez, 2007), and NUCLEOSOME ASSEMBLY PROTEIN1 (NAP1), that transfers H2A-H2B dimers (Galichet and Gruissem, 2006; Zhu et al., 2006) to the newly synthesized DNA.

During G2, the gap between DNA replication and mitosis, DNA integrity is monitored (G2 checkpoint). Under DNA damage situations, WEE1 kinase phosphorylates CDKA producing G2 arrest (De Schutter et al., 2007). A plant specific regulatory DNA repair network has been recently described, involving CDKB1/CYCB1 upregulation in G2 (Weimer and al, 2016). To promote the G2/M transition, a similar cascade to the one in G1/S takes place: the kinases CDKD and CDKF activate CDKA/CYC (Inze and De Veylder, 2006) and CDKB/CYC (Boudolf et al., 2004) to promote the phosphorylation of a variety of targets needed to enter mitosis. The H3

histone variant CENH3, homologous to animal CENP-A, and typical of centromere regions is deposited in late G2 (Lermontova et al., 2006) prior to chromosome condensation.

In mitosis the chromosomes are condensed during prophase. Several phosphorylations on the tail of H3 histones promote compaction, in particular those in the Serine 10 made by AURORA kinases (Demidov et al., 2005; Kawabe et al., 2005). Chromosome condensation is accompanied of a general transcription shutdown. Then, condensed chromosomes move to the metaphasic plate where chromatids are segregated in anaphase to opposite poles of the newborn cells. At the metaphase to anaphase transition, CYCB is degraded by a specific ubiquitin E3-ligase known as ANAPHASE PROMOTING COMPLEX (APC; (Weingartner et al., 2004; Weingartner et al., 2003). APC activity is dependent on three CDH1-like subunits, CELL CYCLE SWITCH 52 (CCS52; (Fülöp et al., 2005; Tarayre et al., 2004)). Finally, during cytokinesis, the phragmoplast will form the cell plate that divides the daughter cells (Boruc and Van Damme, 2015; Gutierrez, 2009; Muller and Jurgens, 2016).

The endocycle is a widespread variation of the cell cycle in which cells replicate their DNA without the following chromosomal segregation during mitosis, leading to an increment of the ploidy level (Fig. 3B;(Edgar et al., 2014)). It is present in eukaryotic organisms such as Drosophila, mouse, humans or plant cells (Edgar et al., 2014). In plants, the endocycle is coupled to differentiation in many cell types. In Arabidopsis roots, the endocycle program starts at the transition domain of the RAM prior to the rapid growth of the cells in the elongation zone (Fig. 2B; (Breuer et al., 2010; Hayashi et al., 2013)). The mitotic to endocycle switch is achieved by the suppression of mitotic-CDK (M-CDK) activity. In Arabidopsis, the APC/C and other tissue specific ubiquitin ligases control endocycle initiation by degrading B-type cyclins distinctive of the G2/M transition (Edgar et al., 2014). For instance, at the boundary between the RAM and the elongation zone, where cytokinins are expressed, they trigger the specific expression of CCS52A1 protein (Boudolf et al., 2009; Vanstraelen et al., 2009). Also, the plant specific CDK inhibitor SIAMESE (SIM) blocks G2 activity of the plant specific CDKB1 (Churchman et al., 2006). Endocycle progression is achieved by an oscillation between high levels of CDKA and inhibition through KRPs. CDKA inhibition allows a G1-like phase where ORIs are thought to be licensed to carry out DNA replication during the endocycle S-phase, although the exact mechanism and regulation of licensing in endocycling cells is still unknown. G/S transition is achieved by the RBR dependent inhibition of the SCF^{FBL17} (SKP1/CULLIN/F-Box FBL17) ubiquitin ligase that degrades KRP, via CDKA activity (Edgar et al., 2014; Gusti et al., 2009; Zhao et al., 2012).

1.4 Licensing of Origins of DNA Replication

The ORIGIN RECOGNITION COMPLEX (ORC) marks, at the end of mitosis, all the potential sites where DNA replication may start (ORIs). However, only a subset of the putative ORIs will activate in the so-called ORI decision point (Wu and Gilbert, 1996). During late mitosis and G1, ORC will recruit other proteins to form the pre-replication complex (pre-RC). This process is called licensing of the ORIs. The proteins that form the pre-RCs are highly structurally and functionally conserved among eukaryotes. ORC is a heterohexameric complex (ORC1 to ORC6) that form an ATPase that provides the necessary energy for the loading of the rest of the pre-RC components, namely, the ATPase CDC6 (CELL DIVISION CYCLE6), CDT1 (CDC10 DEPENDENT TRANSCRIPTION1) and the helicases MCM (MINICHROMOSOME MAINTENANCE, MCM2 to MCM7; Fig. 4; (Masai et al., 2010)). Although the proteins that form the pre-RCs are quite conserved, the regulation is specific of each organism. For example, while in yeast and mouse cells, after ORC and CDC6 loading, CDT1 forms a complex with MCMs prior to the binding of the rest of the pre-RC (Remus et al., 2009; Tanaka and Diffley, 2002; You and Masai, 2008), in *Xenopus laevis* egg extracts and human cells this is not the case (Fragkos et al., 2015; Maiorano et al., 2000). In Arabidopsis, the genes encoding several components of the pre-RC are duplicated, being the case for *ORC1*, *CDC6* and *CDT1*. Although some of them have differential expression domains, for instance *ORC1* (Diaz-Trivino et al., 2005), functional studies of *CDT1a* and *CDT1b* probed partial functional redundancy (Domenichini et al., 2012).

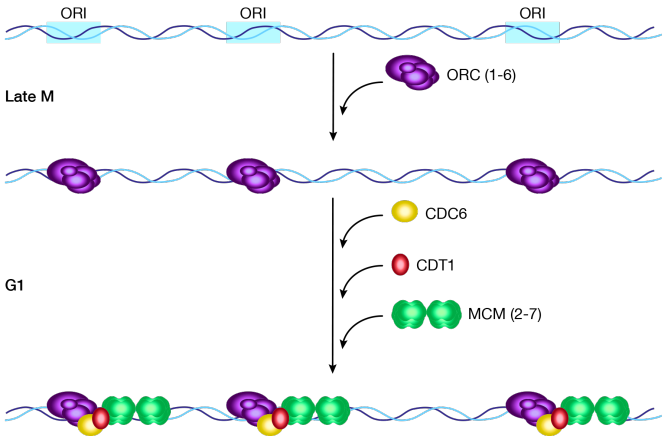


Figure 4 – Origin licensing. Licensing starts at the end of mitosis when the Origin Recognition Complex (ORC) binds to all the potential sites where DNA replication may initiate. During G1 the components of the pre-replication complex (pre-RC) are recruited, CDC6, CDT1 and two MCM helicases.

In humans, mutations in the pre-RC proteins and especially in *ORC1* are associated with Meier-Gorlin syndrome (MGS). This condition considered a form of primordial dwarfism is primarily characterized by intrauterine growth retardation, short stature in the adulthood and underdeveloped kneecaps, ears and head size (Bicknell et al., 2011; de Munnik et al., 2012). Deficiency in pre-RC would be translated into poor ORI licensing, problems in DNA replication

and possible genomic instability that ultimately lead to MGS, immunodeficiency or other rare diseases (Munoz and Mendez, 2017).

ORC was first purified from *Saccharomyces cerevisiae*, and the six subunits were numbered from the largest to the smallest (Bell and Stillman, 1992). ORC1 to ORC5 contain AAA^+ ATPase domains with Walker A and B ATP binding motifs, and winged-helix (WH) domains at the C-termini (Neuwald et al., 1999; Speck et al., 2005). ORC1 possesses, in most eukaryotes studied, a bromo-adjacent homology (BAH) domain at the N-terminal (Bell et al., 1995), related to the BAH domain of yeast Sir3 (Hickman and Rusche, 2010). In plants, an extra domain is present in the N-terminal, a protein-binding plant homeodomain (PHD). CDC6 also contains an AAA^+ ATPase and a WH domain related to the ones found in ORC1 (Duncker et al., 2009). The pre-RC components in archaeal organisms, although simpler are evolutionary conserved with those in eukaryotes (Barry and Bell, 2006; Robinson and Bell, 2005). Remarkably, the ORC1/CDC6 protein, a homologue of both eukaryotic ORC1 and CDC6 proteins, recognizes archaeal ORIs (Ausiannikava and Allers, 2017). Archeas encode at least two ORC1/CDC6 proteins and depending on the species they would recognize the ORI as a monomer (Gaudier et al., 2007) or as a heteromeric complex (Dueber et al., 2007). In protists, there is a high heterogeneity in the composition and the subunits of the ORC complex. While *Plasmodium falciparum* or *Tetrahymena thermophila* present an ORC1 similar to that in yeast (Donti et al., 2009; Gupta et al., 2009; Gupta et al., 2008; Mohammad et al., 2007), in the kinetoplastids *Trypanosoma brucei* and *Leishmania major* an ORC1/CDC6 protein was found, with structural similarities to the ones of Archaea (Godoy et al., 2009; Kumar et al., 2008). All these data suggest a common ancestor for the ORC1 and CDC6 proteins.

Recently, the structure of the Drosophila ORC core complex has been resolved (Bleichert et al., 2015). ORC1 to ORC5 interact through their WH domains forming a ring at the top of the ATPase ring, formed by the union of the five AAA^+ domains. These two rings form a channel in the center of the complex where DNA binding elements encircle the DNA molecule. After MCM helicase loading, ORC1 undergoes a conformational change that seal the central channel, suggesting two ORC conformations that could control cell cycle and developmental ORC functions. The BAH domain of the large subunit of the complex, ORC1, is responsible for chromatin recognition. It also serves as an anchoring site to stabilize the complex onto chromatin. The structure of the mouse ORC1-BAH (Kuo et al., 2012) revealed a cage that specifically recognizes histone H4 dimethylated at lysine 20 (H4K20me2). In animals, the H4K20me status oscillates during cell cycle and its misregulation leads to DNA damage and cell cycle defects (Oda et al., 2009; Schotta et al., 2008). MGS mutations in the mouse ORC1-BAH disrupt the binding of the protein to the histone modification leading to ORI licensing defects

(Kuo et al., 2012). In plants, the role of H4K20 during cell cycle has not yet been described, although opposite to the situation in mammals, the trimethylated form (H4K20me3) has been related to euchromatin (Naumann et al., 2005). This suggests H4K20 regulation may have different roles in plants and mammals. In addition, plant ORC1 PHD and BAH domains bind to histone H3. Although initial studies (de la Paz Sanchez and Gutierrez, 2009) pointed out to a specific recognition of the histone H3 trimethylated at lysine 4 (H3K4me3), the crystal structure of the BAH-PHD domains identified a cage for the recognition of the unmodified histone H3 (Li et al., 2016). Arabidopsis ORIs (Costas et al., 2011) are enriched in activating marks such as H3K4me2 and H3K4me3. Whether the recognition of unmodified H3 by ORC1 is an initial step in pre-RC licensing or other factors will stabilize this association is still unclear.

The sequential assembly of the pre-RC during ORI licensing has been thoroughly studied in yeast and recently elucidated. Once ORC and CDC6 are loaded onto chromatin (Speck and Stillman, 2007), CDT1 binds to a single MCM hexamer (MCM2-MC7) to load the helicase onto the ORC-CDC6, forming an ORC-CDC6-CDT1-MCM complex (Evrin et al., 2014; Sun et al., 2013). Loading of the single hexamer of MCM triggers the eviction of CDC6 and CDT1, leading to an ORC-MCM2-7 complex (Ticau et al., 2015). Finally, a second CDC6 and CDT1-MCM are incorporated onto the ORC-MCM complex (Zhai et al., 2017). These observations are in agreement with a model with one ORC per ORI. Nevertheless, other studies point to a two-ORC model for the helicase loading. First, the interactions between the first and second loading of MCMs suggest there are two-ORCs involved (Ticau et al., 2015), and second, the B2 element of the yeast consensus sequences could bind a second ORC complex (Chang et al., 2011) Mutants in *MCM3* that avoid binding between ORC and CDC6 have problems for the first and second MCM loading (Frigola et al., 2013). Whether just one or the two mechanisms are present in ORI licensing has not been established.

1.5 ORI activation at the G1/S transition

DNA replication is a highly regulated cellular process. To ensure that the DNA is copied only once during each cell cycle, the initiation of DNA replication is temporally separated in two processes: ORI licensing and ORI activation (Bell and Labib, 2016). To avoid re-licensing during S-phase that would lead to re-replication problems, in animals CDT1 is sequestered and inhibited by geminin (Wohlschlegel et al., 2000). In plants a sequence orthologue of geminin has not been identified. Nevertheless, in a screen for CDT1 interacting proteins, the plant specific GLABRA2 (GL2) EXPRESSION MODULATOR (GEM) protein was identified (Caro et al., 2007). *GEM* could be a functional homologue in plants for the animal geminin, since plants

lacking the C-terminal half of GEM present increase cell division and changes in cell fate (Caro et al., 2007). However, further analyses are needed to describe a direct effect of Arabidopsis GEM on DNA replication (Caro and Gutierrez, 2007).

CDKs phosphorylate the pre-RC components MCM (Nguyen et al., 2001), ORC (see below), CDC6 (Petersen et al., 1999) and CDT1 (Coulombe et al., 2013; Sugimoto et al., 2004), triggering their inactivation, eviction from the nucleosome or in the case of CDC6 and CDT1, degradation through the ubiquitin proteasome pathway (Petersen et al., 2000; Sugimoto et al., 2008). Additionally, the regulation of the ORC complex is quite different depending on the organism (Fig. 5). In budding and fission yeasts *S. cerevisiae* and *Schizosaccharomyces pombe*, the ORC complex remains bound to the chromatin during the entire cell cycle. However, in S-phase CDKs phosphorylate ORC, leading to the eviction of CDC6 from the pre-RC (Nguyen et al., 2001; Vas et al., 2001). In Drosophila cells, ORC1 is phosphorylated by CDKs during S-phase (Remus et al., 2005) and degraded by the APC proteasome pathway in mitosis (Araki et al., 2005). *X. laevis* ORC will be phosphorylated upon S-phase entry and disengaged from the chromatin (DePamphilis, 2005). In *Caenorhabditis elegans* the subunits ORC1 and ORC2 are excluded from the chromatin and exported out of the nucleus (Sonneville et al., 2012). In human cells ORC1 availability in S-phase is controlled by phosphorylation by CDKs triggering the recognition by the SCF^{Skp2} (SKP1/CULLIN/F-Box Skp2) complex, and its polyubiquitination and degradation (Mendez et al., 2002).

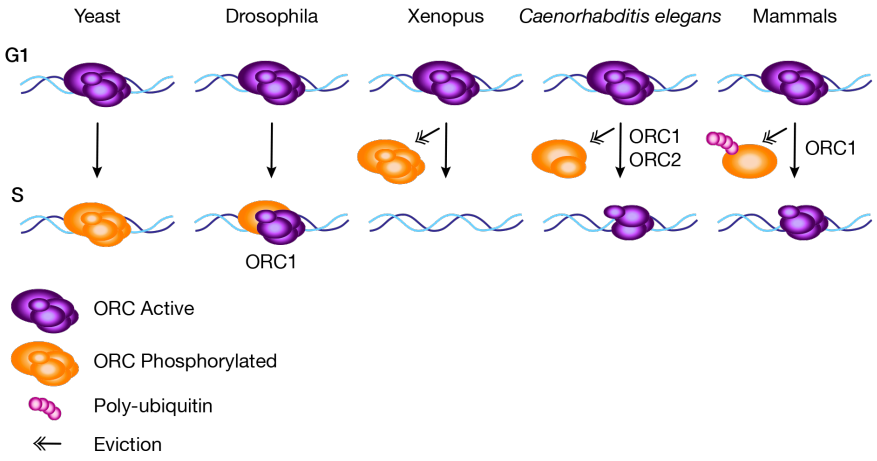


Figure 5 – Regulation of ORC activity. Upon S-phase entry the pre-replication complex activity is inhibited to avoid re-replication. In different organisms the regulation of the ORC complex is different. CDKs phosphorylate the entire complex in yeast or Xenopus cells, leading to the inactivation of the complex and the eviction from the chromatin in the case of Xenopus. In Drosophila, *C. elegans* and mammalian cells, ORC1 is phosphorylated and inactivated, triggering its eviction from the chromatin in *C. elegans*. In mammalian cells, ORC1 is also polyubiquitinated and degraded.

ORI firing in S-phase involves the formation of a pre-initiation complex (pre-IC) and activation of the MCM helicases. In yeasts, Dbf4 dependent kinase (DDK) phosphorylates MCM complexes (Francis et al., 2009; Sheu and Stillman, 2006) facilitating the loading of Sld3 and Sld7 and the replacement of CDC6 by CDC45 (Deegan et al., 2016; Tanaka et al., 2007; Walter, 2000). In a coordinated manner, CDKs phosphorylate Sld3 (Tanaka et al., 2007; Zegerman and Diffley, 2007) to enable loading of Sld2 and GINS (Muramatsu et al., 2010; Yeeles et al., 2015). Once the CMG (CDC45-MCM-GINS) helicase complex is assembled, the pre-initiation complex (pre-IC) is formed, although the helicase is not active yet. CDK phosphorylations trigger the interaction with topoisomerases and MCM10. Once the helicases are active other factors are recruited to the replication bubble, such as DNA polymerase α, RPA (REPLICATION PROTEIN A) or PCNA (PROLIFERATING CELL NUCLEAR ANTIGEN) (Fragkos et al., 2015; Zhai et al., 2017) to initiate DNA synthesis.

1.6 ORI specification

The mechanism of ORI selection depends on the eukaryotic model, reinforcing the need for detailed comparative analysis to fully understand the basic events involved in ORI specification and function. In *S. cerevisiae*, ORIs are structured in autonomous replicating sequences (ARS). The ARS consists of an A-element, which contains the ARS consensus sequence (ACS), and three B-elements B1, B2 and B3 (Marahrens and Stillman, 1992). The ACS is a conserved 11 bp sequence (A/T)TTTA(T/C)(A/G)TTT(A/T), common for all ARSs and essential for origin function (Newlon and Theis, 1993) because it is the ORC recognition site (Bell and Stillman, 1992). The fission yeast *S. pombe* replicator elements comprise DNA sequences ranging between 500 to 1000 bp that lack a well-defined consensus sequence. They are composed of A/T islands specifically recognized by ORC4 (Chuang and Kelly, 1999). Particularly, they were found at intergenic regions colocalizing with promoters (Gomez and Antequera, 1999).

During the last decade, genome-wide mapping of ORIs in diverse organisms has been possible due to the development of techniques such as hybridization of genomic probes (chip) and next generation sequencing (NGS). ORIs detected in *X. laevis* eggs are organized in clusters of regular intervals, lack a consensus sequence, and consist of asymmetric AT tracks (Stanojcic et al., 2008). In contrast, other eukaryotic ORIs are enriched in GC. In mouse cells, ORIs are largely associated with genes, and the most efficient ORIs colocalize with CpG islands at gene promoters (Sequeira-Mendes et al., 2009). They are organized in groups of site-specific but flexible ORIs, enriched in GC nucleotides (Cayrou et al., 2011). Drosophila ORIs share those

characteristics (Cayrou et al., 2011). Although there is no motif of epigenetic mark associated with Drosophila ORIs, mapping studies in different cell types suggest that ORI localization mirrors the cell-type transcriptional program (Comoglio et al., 2015). In human cells, similar to the situation in mouse cells, ORIs correlate with the genomic landscape. Human ORIs associate with GC-rich regions at CpG islands (Cadoret et al., 2008), correlate with G-quadruplex (G4) motifs (Besnard et al., 2012), and replicate early in most of the cases (Picard et al., 2014). The study of different human cell types suggests that cell-type specific ORIs depend on cell-type specific transcriptional programs. Also, two potential key regulators of ORI selection were described H4K20me1 and H3K27me3 (Picard et al., 2014). The last analysis of ORIs in mouse cells, in combination with 43 chromatin marks identified three ORI classes (Cayrou et al., 2015). Low efficient ORI, depleted in chromatin marks form the largest class 1. ORIs enriched in enhancer elements compose class 2, while the most efficient ORIs associated with open chromatin and polycomb regions form the class 3. All three classes coincide with nucleotide depleted and GC-rich regions that may potentially form G4 structures (Cayrou et al., 2015). A recent study of ORIs during *C. elegans* embryo development identifies a reorganization of the ORIs after gastrulation onset (Rodriguez-Martinez et al., 2017). At pre-gastrula stage ORIs associate to open chromatin and after gastrulation new ORIs correlate to enhancers and CpG islands. Thus, ORI selection is coordinated with the specific transcriptional programs during development (Rodriguez-Martinez et al., 2017). Mapping of nucleosome position and ORIs in *L. major* indicates that active transcription is determinant for ORI selection. ORIs correlate with RNA polymerase pausing sites, providing higher flexibility to ORI activation (Lombrana et al., 2016). In Arabidopsis cultured cells, ORIs colocalize with the 5' half of genes, especially of highly expressed genes (Costas et al., 2011). Accordingly, most ORIs are enriched in activating marks such as the histone variant H2A.Z, di-trimethylation at the lysine 4 of histone H3 (H3K4me2/3) or acetylation of histone H4 (H4Kac). Remarkably, ORIs correlate with short GC-rich stretches even though the Arabidopsis genome is rich in AT (63.8%; (Costas et al., 2011)). In general, eukaryote ORI mapping revealed (i) lack of consensus DNA replication sequence, (ii) association to transcription and active chromatin and (iii) association with GC-rich regions that may form DNA structures such as G4.

1.7 Chromatin organization

Chromatin is organized in heterochromatic domains, which are densely compacted for most of the cell cycle and in euchromatin, with a relatively less dense organization. Genomic features are not evenly distributed throughout the chromosomes, as genes are more frequent

in the euchromatic chromosome arms. In Arabidopsis, the constitutive heterochromatic regions are located at the pericentromeric sites, at telomeres, and in the nucleolus organizing regions (Feng and Michaels, 2015; Pecinka et al., 2004; Schubert et al., 2012). In addition, there are silent domains within the euchromatic arms enriched in repressive marks. These regions are composed mainly of transposable elements (TEs), inserted within euchromatic regions, and of Polycomb related genes (Feng and Michaels, 2015; Fransz et al., 2006). Polycomb group proteins repress gene expression programs that are not needed at a given developmental stage or for a particular cell status, for instance during differentiation processes. Effective silencing is achieved by two modifications, namely, trimethylation of lysine 27 of histone H3 (H3K27me3) by the Polycomb Repressive Complex 2 (PRC2) or monoubiquitynation of lysine 121 of histone H2A (H2AK121Ub; lysine 119 in animals) by the Polycomb Repressive Complex 1 (PRC1; (Xiao and Wagner, 2015)).

Recently, different strategies to identify interactions between genomic sites have been developed and used in Arabidopsis. A first conclusion of these studies is that the overall 3D interaction network resembles that of Drosophila and mammalian cells (Sequeira-Mendes and Gutierrez, 2016). This is particularly striking for the separation between euchromatin and heterochromatin (Grob et al., 2014). In addition, Hi-C experiments identified genomic regions that have the general features of active chromatin that establish distal interactions with other similar domains. Short-range interactions also occur between the 5' and the 3' end of genes, particularly in highly expressed genes (Liu et al., 2016). By contrast, genomic domains that have the global properties of repressed chromatin establish contacts with similar regions and are separated from active domains (Feng et al., 2014; Wang et al., 2015). Remarkably, heterochromatic regions interspersed along euchromatic chromosome arms tend to contact each other both in *cis* and in *trans*. This leads to the formation of a specific heterochromatin region, called a KNOT (Grob et al., 2014). Interestingly, other repressed regions that establish long-range interactions have been reported to contain promoters that are enriched in the H3K27me3 Polycomb mark (Liu et al., 2016). This suggests that such interactions might contribute to the coordinated expression of those genes. The formation of genome territories that are well separated in TADs (topologically associating domains), as described for Drosophila and mammalian cells (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), does not seem to be a characteristic of the Arabidopsis genome. The lack of TADs might be a consequence of the lack in plants of a structural homologue of CTCF in mammals and CP190 in Drosophila, the proteins that serve as an insulator that defines TAD boundaries (Barutcu et al., 2017; Gonzalez-Sandoval and Gasser, 2016). Although typical TADs are missing from Arabidopsis, regions with functional similarities have recently been reported (Liu and Weigel,

2015; Wang et al., 2015). Therefore, it could be very interesting to determine how these TAD-like regions are established and whether they are developmentally regulated or respond to hormonal and environmental cues.

Recent works have focused on the analysis of genome-wide maps of dozens of DNA and histone modifications and the different combination of patterns of modifications or distinct chromatin states. In a recent study (Sequeira-Mendes et al., 2014), silent heterochromatin was separated in two types depending on the GC content, AT-rich heterochromatin and GC-rich heterochromatin. Both chromatin states colocalize with constitutive heterochromatin, which is mainly composed of TEs. Arabidopsis constitutive heterochromatin is characterized by high levels of DNA methylation in three sequence contexts (CG, CHG and CHH), dimethylation of the lysine 9 of histone H3 (H3K9me2) and the plant-specific epigenetic mark monomethylation of the lysine 27 of histone H3 (H3K27me1). Different processes have evolved to maintain TE silenced. In Arabidopsis, C methylation results from a combination of the activities of MET1 (METHYLTRANSFERASE 1 that is responsible for mCG; (Lister et al., 2008)), CMT2, CMT3 (CHROMOMETHYLTRANSFERASE 2 and 3, which produce mCHH and mCHG, respectively; (Stroud et al., 2014; Zemach et al., 2013)), and DRM2 (DOMAINS REARRANGED METHYLASE2, generates mCHH) as part of the RNA-dependent DNA methylation (RdDM) pathway (McCue et al., 2015). The RdDM pathway relies on RNA Pol IV dependent 24-nucleotide short interfering RNAs (siRNAS; (Herr et al., 2005; Onodera et al., 2005; Pontier et al., 2005)) and RNA Pol V-dependent RNAs (Wierzbicki et al., 2008) to keep TEs silent across generations (Bohmdorfer et al., 2016; Fultz et al., 2015). Out of the fifteen Arabidopsis putative H3K9 methyltransferases, only three are responsible for the majority of this epigenetic mark: KRYPTONITE (KYP, also known as SUVH4 for SU(VAR)HOMOLOGUE), SUVH5 and SUVH6 (Bernatavichute et al., 2008; Du et al., 2015). A self-reinforcing loop between histone H3K9me2 and DNA methylation has already been reported. Mutants in *KYP* have a reduction in H3K9me2 levels besides low CHG methylation (Jackson et al., 2002). Furthermore, *CMT3* mutation compromises both DNA and histone methylation (Tariq et al., 2003).

The plant specific silencing mark H3K27me1 is deposited by the ARABIDOPSIS TRITHORAX-RELATED PROTEIN 5 (ATXR5) and ATXR6 (Jacob et al., 2009). Loss of function *atxr5 atxr6* mutations leads to a disruption of constitutive heterochromatin and transcriptional activation of TEs without affecting DNA methylation (Jacob et al., 2009). The genomic instability in these mutants produces an excess of DNA corresponding to the deregulated heterochromatic regions (Jacob et al., 2010). Recent studies suggest that the genomic instability is produced by the transcriptional activation of TEs during S-phase (Hale et al., 2016).

There are two classes of TEs that differ in their transposition mechanism: class 1 or retrotransposons and the class 2 of DNA transposons. Even though TEs account for an important fraction of all eukaryote genomes, the particular families that are more prevalent may differ from genome to genome. Thus, whereas in general LTR retrotransposons are the most prevalent type of TEs in plants (Bennetzen and Wang, 2014), another type of retrotransposons, LINEs, are the most prevalent TEs in mammalian genomes (Kazazian, 2000). In Arabidopsis, several TE families account for 21% of the genome and although some of them are scattered along the chromosome arms most TEs concentrate in the pericentromeric heterochromatin (Ahmed et al., 2011; Feng and Michaels, 2015).

2. Aims

DNA replication is a common process in all living organisms. Although the protein machinery is highly conserved among eukaryotes, pre-RC regulation and ORI selection is specific for each species. Cell cycle and DNA replication studies in plants will increase the knowledge of how these important processes have changed during evolution. In addition, analyses in plants enable the possibility to learn about these mechanisms *in vivo* and during organogenesis, for instance using Arabidopsis roots. Arabidopsis also offers the opportunity to study the function of newly duplicated genes. Furthermore, its small genome, already sequenced and well annotated facilitates genomic studies at specific chromatin domains such as the heterochromatic regions. Therefore, the major aim of this work is to investigate the control of DNA replication using Arabidopsis culture cells and Arabidopsis roots as model systems. To achieve this goal we propose to address:

1. The regulation of the two Arabidopsis ORC1 proteins during root organogenesis.

2. The dynamics of ORC1 proteins during cell cycle and endocycle processes.

3. The characteristics that define active ORIs in the heterochromatin of Arabidopsis cells.

3. Materials and Methods

3.1 Materials

Bacterial strains

The cloning of *ORC1* genes was carried out with *Escherichia coli* DH5α. GATEWAY empty vectors were maintained using *E. coli* DB3.1 strain. *Arabidopsis thaliana* plants were transformed employing *Agrobacterium tumefaciens* C58C1 strain. All bacteria were grown in LB medium supplemented with the necessary antibiotics at 30 or 37 °C for *A. tumefaciens* and *E. coli*, respectively.

Plant growth conditions

A. thaliana MM2d suspension cell line (Menges and Murray, 2002) was grown in the absence of light at 26 °C with agitation. The cells were subcultured every 7 days into fresh Murashige & Skoog medium (MS, pH 5.8, Duchefa) supplemented with 3% sucrose (Duchefa), 0.5 µg/mL 1-naphthaleneacetic acid (Duchefa), 0.1 µg/mL kinetin (Sigma), and 0.103 µg/mL vitamins (Duchefa).

A. thaliana plants (Columbia ecotype, Col-0) were grown in 0.5x MS medium (pH 5.7) supplemented with MES (Sigma), vitamins (Duchefa) 0.5 or 1% sucrose (Duchefa), and 0.8 or 1% agar (Duchefa). Plants grew in an incubator at 20 °C and 60% moisture, under long day conditions (16 h light, 8 h dark, fluorescent tubes Philips MASTER TLD Super80, 36 W, 4000 K, 100 µmol/m^2/s). MS medium was supplemented with antibiotics for plant selection and specific drugs according to the treatments.

Transgenic plants generated

pORC1a(-1324)::ORC1a-GUS/G3GFP
pORC1a(-87)::ORC1a-GUS/G3GFP
pORC1b(-793)::ORC1b-GUS/G3GFP/mRFP

T-DNA insertion lines characterized

orc1a-1: WiscDsLox287F12
orc1b-1: SALK_042536C
These lines were obtained from the Arabidopsis Biological Resource Center (ABRC).

Other plant lines used in this study

pCYCB1;1::GFP (Colon-Carmona et al., 1999; Ubeda-Tomas et al., 2009)
pFBL17::FBL17-GFP (Noir et al., 2015)
pCDT1a::CDT1a-GFP (Lopez, unpublished)
pORC6::ORC6-GFP (Diaz-Trivino, 2005)
pORC2::ORC2-GFP (Ngo et al., 2012)
pMCM7::MCM7-GFP (Herridge et al., 2014)
fbl17-1 (Gusti et al., 2009)
skp2a-1 (Ren et al., 2008)
atxr5 (Jacob et al., 2009)
atxr6 (Jacob et al., 2009)

3.2 Molecular biology techniques

Molecular cloning

The genomic fragment of *ORC1a* (AT4G14700) gene, containing the promoter and coding region, and excluding the termination codon and the 3'UTR, was amplified from F4C24 BAC (Mozo et al., 1998). The amplification was done by touchdown PCR (Korbie and Mattick, 2008) using Accuprime Taq DNA polymerase High Fidelity (Invitrogen). The genomic region of *ORC1b* (AT4G12620) gene, containing the promoter and coding region, and excluding the termination codon and the 3'UTR, was amplified from genomic DNA by PCR using Platinum Pfx DNA polymerase (Invitrogen). The primers used for the amplification of these genes are listed in Supplementary Table 1. The PCR products were purified using Wizard SV Gel and PCR Clean-Up System (Promega), cloned into pDONR221 (Invitrogen) using BP Clonase II (Invitrogen) and amplified in *E. coli* DH5α. The resulting pENTRY vectors were subjected to restriction enzyme analysis and sequencing to discriminate those without mutations. Selected clones were recombined into the following Gateway Destination vectors (Nakagawa et al., 2007) using LR Clonase II (Invitrogen):

pGWB433: allows a C-terminal fusion of the protein to GUS and provides plants with resistance to kanamycin. Used for *ORC1b* cloning.

pGWB450: allows a C-terminal fusion of the protein to G3GFP and provides plants with resistance to kanamycin. Used for *ORC1b* cloning.

pGWB453: allows a C-terminal fusion of the protein to mRFP and provides plants with resistance to kanamycin. Used for *ORC1b* cloning.

pGWB533: allows a C-terminal fusion of the protein to GUS and provides plants with resistance to hygromycin. Used for *ORC1a* cloning.

pGWB550: allows a C-terminal fusion of the protein to G3GFP and provides plants with resistance to hygromycin. Used for *ORC1a* cloning.

Expression vectors were amplified in *E. coli* DH5α and checked by restriction enzyme analysis. F4C24 BAC was obtained from the ABRC and Tsuyoshi Nakagawa from the Research Institute of Molecular Genetics (Japan) provided the destination vectors.

Extraction of genomic DNA

Either rosette leaves or complete seedlings were frozen in dry ice and ground with glass beads using a Silamat S5 device (Ivoclar Vivadent) for 10 s. The samples were incubated with 200 µL per leaf of extraction buffer (140 mM D-Sorbitol, 220 mM Tris-HCl, pH 8.0, 22 mM EDTA

pH 8.0, 800 mM NaCl, 0.8% CTAB, 1% N-Lauroylsarcosine) for 10 min at 65 °C with mild agitation. The lysate was mixed with an equal volume of chloroform and centrifuged at 18,000x*g* for 10 min. The aqueous phase was transferred to a fresh eppendorf tube and DNA was precipitated with one volume of isopropanol for 10 min. The DNA was pelleted by centrifugation at 18,000x*g* for 10 min, washed with 70% ethanol and air-dried before resuspension in water.

Genotyping PCRs

To select homozygous T-DNA insertion mutant plants, total genomic DNA was analyzed by PCR using Taq DNA polymerase (Biotools) and the primers listed in Supplementary Table 2.

Purification of short replication intermediates

The short nascent strands (SNS) from replication intermediates, used in the ORI activity assays, were purified essentially as described in (Costas et al., 2011). At day 4 after passage, 100 mL of the asynchronous cell suspension were either directly collected for SNS preparation or synchronized at the desired time points (2, 3.5 or 7 h) before SNS isolation. Cells were collected by filtration, frozen and ground in liquid nitrogen. Total DNA was purified in RNase-free conditions. Ground cells were lysed in SONI buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, pH 8.0, 2% SDS, 100 mM LiCl and 100 µg/mL proteinase K). Total genomic DNA was extracted twice with phenol:chloroform:isoamylalcohol (25:24:1), precipitated with 2 volumes of ice-cold ethanol in the presence of sodium acetate, washed with 70% ethanol and air-dried before resuspension in TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0).

Purified DNA was denatured by heating 10 min at 100 °C and size-fractionated in a seven-step neutral sucrose gradient (5-20% sucrose in TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0 and 100 mM NaCl)), by centrifugation at 102,300xg in a SW-40T1 Beckman rotor for 20 h at 20 °C (Gomez and Antequera, 2008). Fractions (1 mL) were collected from the top and the DNA was ethanol-precipitated. An aliquot of each fraction was analyzed in a 1% alkaline agarose gel (50 mM NaOH, 1 mM EDTA) to monitor size fractionation (Fig. 6A). Fractions containing replication intermediates ranging between 300-2000 nt in size were treated with 0.5 U/µL of polynucleotide kinase (PNK, Fermentas) to phosphorylate 5'hydroxyl ends in the presence of 1 mM dATP for 30 min at 37 °C. After PNK inactivation, phosphorylated DNA was extracted, precipitated and resuspended in water. λ-exonuclease degrades contaminating random sheared DNA leaving untouched DNA replication intermediates that are protected by a 5'RNA-primer (Gerbi and Bielinsky, 1997). The λ-exonuclease digestion was

carried out with 0.5 U/µL of enzyme (New England Biolabs) following manufacture's instructions at 37 °C overnight. The efficiency of the digestion was monitored by adding 40 ng of phosphorylated linearized plasmid to an aliquot of each reaction tube (Fig. 6B). DNA from each λ-treated fraction was extracted, precipitated and resuspended in TE. The relative abundance of nascent DNA strands around putative origins was monitored by qPCR using primer sequences listed in Supplementary Table 3.

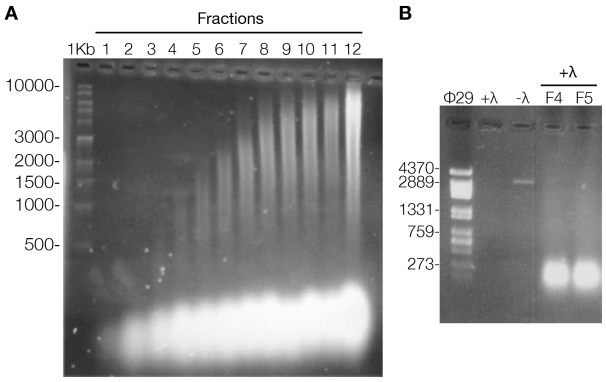


Figure 6 – Purification of short DNA nascent strands. (A) Fraction profile of denatured total genomic DNA from MM2d cells in a neutral sucrose gradient control in a 1% alkaline agarose gel. 1 Kb: size markers in bp. **(B)** λ-exonuclease digestion controls. Fractions containing replication intermediates ranging between 300-2000 nt (F4 and F5) in size were subjected to digestion by λ-exonuclease (+λ). A linearized and phosphorylated plasmid (~3 Kb; observed in –λ) was co-digested with an aliquot of each fraction to monitor the efficiency of the reaction. Φ29: size marker in bp.

Isolation of RNA

Either MM2d cells or seedlings were frozen together with glass beads in liquid nitrogen. The material was ground for 8 s using a Silamat S5 device (Ivoclar Vivadent). 500 µL of Trizol reagent (Invitrogen) were added per 100 mg of material and ground again for 8 s. Another 500 µL of Trizol were added, thoroughly mixed and incubated for 5 min at room temperature. 200 µL of chloroform per mL of Trizol were used to extract total RNA by centrifugation at 20,000x*g* for 10 min at 4 °C. The upper phase was transferred to a fresh tube and the RNA was precipitated with 20 µg of glycogen (Roche) and 500 µL of ice-cold isopropanol per 1 mL of Trizol for 10 min at room temperature. RNA was pelleted by centrifugation at 20,000x*g* for 20 min at 4 °C, washed with 75% ethanol, air-dried and resuspended in nuclease free water. Contaminating DNA was digested with 0.1U/µL of DNase I (Roche) for 20 min at 37 °C. After heat inactivation of the enzyme, RNA was extracted with phenol:chloroform:isoamylalcohol (25:24:1) using heavy phase lock gel tubes (3 Prime), precipitated with 20 µg of glycogen (Roche), 1/10 3 M sodium acetate, pH 5.2 and 2.5 volumes of ice-cold ethanol, washed with 75% ethanol, air-dried and resuspended in nuclease free water. Total RNA was quantified using Nanodrop (Thermo Scientific) and its integrity was assessed by agarose gel electrophoresis (Fig. 7).

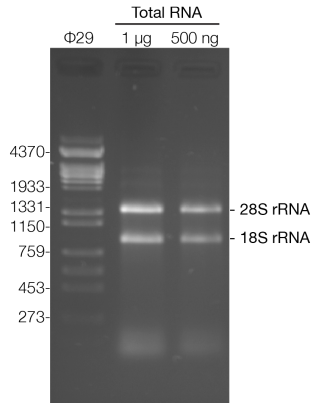


Figure 7 – Evaluation of RNA integrity by agarose gel analysis. Samples of total RNA (1 µg or 500 mg) were fractionated on a 1% agarose gel. The integrity was considered correct when the bands corresponding to 28S rRNA and 18S rRNA were detected. Φ29: size marker in bp.

One microgram of the isolated RNA was reverse-transcribed with SuperScript III (Invitrogen) using either oligo-dT primer (mRNA) or random hexamers (total RNA). A reaction without reverse transcriptase (-RT) was included to verify the absence of genomic DNA in the sample. 1 µL of a 3-fold diluted cDNA reaction were used as template in PCR using the primer sequences listed in Supplementary Table 4. 2 µL of a 3-fold diluted cDNA reaction were used as template in qPCR using primer sequences listed in Supplementary Table 3.

Chromatin immunoprecipitation (ChIP)

ChIP was performed as described in (Desvoyes, 2017). 50 mL of MM2d cells were harvested 4 days after subculture and fixed using ice-cold 1% formaldehyde in PBS by applying vacuum infiltration (3 rounds of 6 min on/4 min off, 85,000 Pa). The cross-linking was stopped by the addition of 0.125 M glycine, and vacuum infiltration for another 5 min. Cells were washed with ice-cold water, collected by filtration, frozen and ground in liquid nitrogen. The ground material was resuspended in 30 mL of Extraction Buffer (0.25 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail for plant cell extracts (Sigma)). Nuclei were released using a dounce homogenizer device with a tight pestle (0.05 ± 0.025 mm clearance) and a loose pestle (0.114 ± 0.025 mm clearance), and filtered through a double Miracloth mesh. After centrifugation of the sample for 20 min at 3,000x*g* and 4 °C, nuclei were resuspended in 2 mL of Lysis Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1x protease inhibitor cocktail) and disrupted by sonication in a Bioruptor Plus (Diagenode) for 30-45 cycles of 30 s on and 30 s off, at high power mode. Soluble chromatin was separated by centrifugation at 20,000x*g* for 5 min at 4 °C.

The sonication efficiency was tested in an aliquot of the soluble chromatin. First, cross-links were reverted with 200 mM NaCl and an overnight incubation at 65 °C with strong agitation. Then, RNA was digested with 40 µg/mL of RNase A and 0.1 U/µL RNase T1 (RNase A/T1 mix, Thermo Scientific) for 1 h at 37 °C and mild agitation. The digestion of proteins was

carried out with 50 µg/mL of proteinase K for 2 h at 37 °C and mild agitation. After that, the DNA was extracted with phenol:chloroform:isoamylalcohol (25:24:1) using heavy phase lock gel tubes (3 Prime), precipitated with 20 µg of glycogen (Roche), 1/10 3 M sodium acetate, pH 5.2 and 2.5 volumes of ice-cold ethanol, washed with 75% ethanol, air-dried and resuspended in nuclease free water. The amount of DNA was quantified using the Qubit system and the Qubit dsDNA High Sensitivity assay kit (Life technologies). Finally 1 µg of sonicated DNA was evaluated on a 1% agarose gel. Samples ranging between 200-700 nt were used for immunoprecipitation.

Soluble chromatin was diluted ten times with ChIP dilution buffer (16.7 mM Tris-HCl, pH 8.0, 167 mM NaCl, 1.2 mM EDTA, 1.1% Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail) to decrease SDS concentration to 0.1%. To eliminate unspecific background samples were pre-cleared for 1h with 30 µL of protein G Plus agarose beads (Santa Cruz Biotechnology). After that, 100 ng of pre-cleared chromatin was taken as input and 1 µg was employed per ChIP reaction, using the following antibodies: anti-H3K9me2 (Abcam ab1220, 3 µg), anti-H3K27me1 (Millipore 07-448, 1 µg), anti-total H3 (Abcam ab1791, 2 µg), or anti-rat IgG (Abcam ab6703, 2 µg) as a negative control. The ChIP reactions were incubated overnight in a rotating wheel at 4 °C. Immune complexes were recovered with 50 µL of protein G Plus agarose beads for 2 h in a rotating wheel at 4 °C. Beads were washed with four different buffers, twice with 1 mL each, first a quick wash after mixing by inversion and then incubating for 5 min in a rotating wheel at 4 °C. Beads were spun down by centrifugation at 1,000x*g* for 2 min at 4 °C. First the samples were washed with low salt wash buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 150 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail), then with high salt wash buffer (20 mM Tris-HCl, pH 8.0, 2 mM EDTA, 500 mM NaCl, 0.1% SDS, 1% Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail), LiCl wash buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.25 M LiCl, 1% Igepal CA-639, 1% sodium deoxycholate, 1 mM PMSF, 1x protease inhibitor cocktail) and finally TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0). To elute the immune complexes, 200 µL of elution buffer (1% SDS, 0.1 M sodium bicarbonate, pre-warmed at 65 °C) were added to the beads and incubated for 15 min at 65 °C with mild agitation. Beads were spun down, the supernatant transferred to a new tube and the elution step was repeated. Input volume was adjusted to 400 µL with elution buffer.

ChIP and input samples were purified as follows. First, cross-links were reverted with 200 mM NaCl and an overnight incubation at 65 °C with strong agitation. Then, RNA was digested with 40 µg/mL of RNase A and 0.1 U/µL RNase T1 (RNase A/T1 mix, Thermo Scientific) for 1 h at 37 °C and mild agitation. The digestion of proteins was carried out with 50 µg/mL of proteinase K for 2 h at 37 °C and mild agitation. After that, the DNA was extracted with

phenol:chloroform:isoamylalcohol (25:24:1) using heavy phase lock gel tubes (3 Prime), precipitated with 20 µg of glycogen (Roche), 1/10 3 M sodium acetate, pH 5.2 and 2.5 volumes of ice-cold ethanol. DNA was washed with 75% ethanol, air-dried and resuspended in 100 µL of TE. ChIP results were analyzed by qPCR using primer sequences listed in Supplementary Table 3.

Quantitative real-time PCR

All qPCRs (SNS, cDNA, and ChIP) were performed using GoTaq Master Mix (Promega) according to the manufacturer's instructions in an ABI Prism 7900HT machine (Applied Biosystems) using the primers listed in Supplementary Table 3. In each case, the quantification was determined using a standard curve (five serial four-fold dilutions of gDNA or cDNA when possible). SNS enrichment was normalized against a region flanking the ORI under analysis or a region lacking ORIs (negative control). RNA expression levels were compared to the reference gene *GAPC2* (*GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2*, (Wang et al., 2014)). The amount of immunoprecipitated material was estimated as percentage of input chromatin and then normalized against total H3 content.

Isolation of nuclear proteins

To evaluate the expression level of ORC1 proteins, 1 g of 7 dps seedlings was frozen in liquid nitrogen. The material was ground to a fine powder in liquid nitrogen using a mortar and a pestle. The pulverized tissue was resuspended in 10 mL of Extraction Buffer (0.25 M sucrose, 10 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$, 1% Triton X-100, 1 mM PMSF, 1x protease inhibitor cocktail for plant cell extracts (Sigma), 100 µM MG132 (Peptide Institute), 0.5 µM epoxomicin (Peptide Institute)). Nuclei were released using a dounce homogenizer device with a tight pestle (0.05 ± 0.025 mm clearance) and a loose pestle (0.114 ± 0.025 mm clearance), and filtered through a double Miracloth mesh. After centrifugation of the sample for 20 min at 3,000x*g* and 4 °C, nuclei were resuspended in 200 µL of Lysis Buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, 1 mM PMSF, 1x protease inhibitor cocktail, 100 µM MG132, 0.5 µM epoxomicin) and DNA was disrupted by sonication in a MSE 150 watt ultrasonic disintegrator three times, 10 s each. Soluble proteins were separated by centrifugation at 20,000x*g* for 10 min at 4 °C. and quantified using BCA protein assay (Pierce BCA Protein Assay kit, Thermo Scientific). 100 µg of nuclear protein was loaded in an 8% Tris-glycine polyacrylamide gels to run SDS-PAGE electrophoresis and subsequent Western Blot. Briefly, proteins in the gel were transferred to a methanol-activated Immobilon-P membrane (0.45 µm, Millipore) for 90 min at

250 mA using the electro blotting system from Bio-Rad. The membrane was then blocked for 30 min with 5% non-fat milk in PBST (1x PBS, 0.1% Tween-20) and then incubated with the primary antibody overnight at 4 °C (anti-ORC1b diluted 1:3,000). After three washes with PBST the membrane wash incubated with the secondary antibody for 1 h (Amersham ECL rabbit IgG HRP linked (GE Healthcare Life Science) diluted 1:10,000), washed again 3 times and proteins were detected using the kit Immobilon WB Chemiluminescent for HRP substrates (Millipore).

Custom rabbit monoespecific antibody production

The specific antibody against the protein ORC1b was obtained from BioGenes. The ORC1b protein sequence was analyzed, focusing on unique residues not present in ORC1a and basic aminoacids that could be exposed with a higher probability. The selected peptide was GMNLIRKRERAPR from residue 72 to 85. Prior to the immunization step, the pre-immune sera from 16 rabbits were tested against total nuclear protein from wild type Col-0 plants by WB. The two pre-immune sera with less unspecific background signal, especially in the size of the endogenous ORC1b protein (around 100 KDa), were selected and the rabbits immunized using the synthetic peptide. The immunized rabbits produced polyclonal antibodies against the peptide sequence. Serum was collected and tested against total nuclear protein from wild type Col-0 and ORC1b-GFP expressing plants by WB. To determine the specifity of the bands obtained, an immunizing peptide blocking assay was carried out. Before blotting, the antiserum was incubated with an excess of the custom peptide, thus specific antibodies are no longer available to bind to the proteins on the WB, resulting in loss of the specific band. The selected serum was affinity purified against the peptide sequence to isolate only those antibodies in the serum that specifically bind to epitopes contained within the peptide sequence.

3.3 Cell biology techniques

Cell synchronization

Cells in exponential phase (4 days after subculture) were synchronized in G0/G1 by growing them in MS without sucrose for 24 h. To release the cell cycle block the medium was replaced with MS with sucrose (Menges and Murray, 2002). Samples for analysis were taken at 2 (G1/S transition), 3.5 (early S) and 7 (late S) hours. Cell cycle synchronization was monitored by assessing the mRNA levels of *CYCA3;1* and *CYCB1;4* genes, which accumulate in the middle of S-phase and in G2/M, by qPCR according to primers listed on Supplementary Table 3 (Fig. 8).

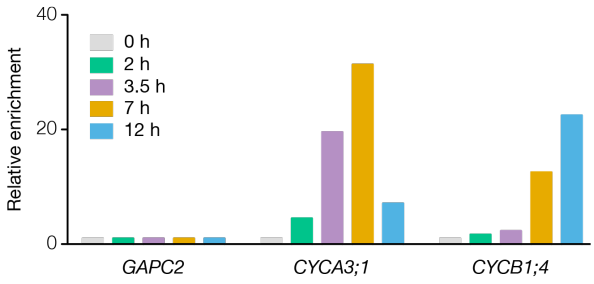


Figure 8 – Expression levels of cell cycle marker genes during MM2d cell synchronization. The RNA level of two well-known cell cycle markers, *CYCA3;1* and *CYCB1;4* was assessed for each synchronization time. Expression was normalized against the reference gene *GAPC-2* (*GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2*).

Transgenic plant generation

Agrobacterium tumefaciens (C58C1 strain) was transformed with the expression vectors and grown for 2 days at 30 °C in plates containing 20 µg/mL rifampicin and 100 µg/mL spectinomycin. A starting bacterial culture of 5 mL was set the day before transformation and 1 mL was used to check the colonies through PCR. Then, a larger culture was initiated from a 2:100 dilution of the starter one and grown for 8 h. Cells were harvested by centrifugation at 3,000x*g* for 20 min and resuspended in a solution of 5% sucrose and 0.05% Silwet L-77. 3-week old Columbia wild type plants were transformed using the floral dip method (Clough and Bent, 1998). In the T1 generation, transformant seeds were selected in 0.8% agar MS plates containing either 15 µg/mL hygromycin (*ORC1a* selection) or 50 µg/mL kanamycin (*ORC1b* selection). In the next generation, the segregation of the resistance trait was followed in at least 50 plants to select only the lines with one insertion. Finally, in the T3 generation, the homozygous lines were selected, as those with all the plants tested (around 30) resistant to the antibiotic.

Histochemical detection of GUS activity

Detection of GUS activity was performed using 5-bromo-4-chloro-3-indoyl-β-D-glucuronide (X-Gluc) as described in (Jefferson et al., 1987). Briefly, 4, 7 or 12 dps seedlings were infiltrated into fresh GUS substrate (5 mM phosphate buffer, pH 7.0, 1 mM potassium ferrocyanide, 1 mM potassium ferricyanide, 0.25% Triton X-100, 5 µg/mL X-Gluc) for 5 min (30,000 Pa) and then incubated for 72 h at 37 °C. Seedlings were removed from the substrate, washed twice with water and preserved in 1x PBS-50% glycerol at 4 °C until observation under the microscope Axioskop2 plus (Zeiss) or the stereomicroscope MZ9.5 (Leica).

Whole root mounting for confocal microscopy

Roots were stained either with 50 µg/mL propidium iodide (Sigma; for cell walls) or 10 µM FM4-64 (Life technologies; for membranes) and directly examined using confocal LSM510 or LSM710 (Zeiss). GFP was detected with 488 nm laser and mRFP with the line 561 nm.

Live imaging

Seedlings were grown for 3 days and then transferred to P35 glass bottom dishes (MatTek) with 100 µL of water. A piece of 1% agar MS solid media was cut and placed on top of the root, covering the seedlings completely except for the aerial part. The dishes were hung vertically in the plant culture chamber for 24 h. Prior to confocal observation, 10 µM FM4-64 or 50 µg/mL propidium iodide were injected under the solid media to stain the membranes or the cell wall, respectively. *In vivo* images were acquired every 1, 2 or 4 min to observe rapid processes or every 15 min to supervise G1 progression, either manually or using the time lapse module. To avoid photobleaching, mild laser exposure conditions were used and the fluorescent intensity of the scans was monitored using the LSM software.

Plant treatments

To study the regulation of the ORC1 proteins, GFP-tagged seedlings were incubated in liquid media with different inhibitors. At least 18 h before confocal observation, 4dps seedlings were transferred into 6-well plates with liquid MS. The plates were placed again in the plant culture chamber. This allowed the roots to recover from the temporary hypoxia produced by the liquid, and therefore GFP recuperates its fluorescence capacity. To inhibit the proteasome pathway plants were incubated with 100 µM MG132 (Peptide Institute) and 0.5 µM epoxomicin (Peptide Institute) for 4 h or with 25 µM MLN4924 (APExBIO) for 6 h. To inhibit the CDK dependent phosphorylation, seedlings were incubated with 50 µM roscovitin (Sigma) for either 4 or 8 h. All these drugs are diluted in DMSO, therefore, control plants were incubated with 0.5% DMSO for the same amount of time. All incubations were carried out in the plant culture chamber.

To assess the root growth in the presence of DNA replication stress conditions, 48 h stratified *ORC1* mutant and Col-0 seeds were plated in 0.5% sucrose 1% agar MS plates. 72 h later, seedlings were transferred to 0.5% sucrose 1% agar MS plates containing either 6.7 µM zeocin (Invitrogen), 1, 2 or 5 mM hydroxyurea (Sigma), 0.12 µg/mL aphidicolin (Sigma) or no drug as a control. Seeds that did not germinate were removed from the analysis. Root growth was measured every 24 h for a total of 10 days.

Flow cytometry

MM2d cells were collected at either 4 or 7 days after subculture by vacuum filtration. Seeds from *ORC1* mutant and Col-0 plants were grown for 7 days. To enrich the sample in root apical meristems, root tips (~5 mm) were cut. MM2d cell retentate or root tips were chopped in cold Galbraith solution (20 mM MOPS, pH 7.0, 45 mM $MgCl_2$, 30 mM sodium citrate, 0.1% Triton X-100, pH 7.0, (Galbraith et al., 1991)) using a single edge razor blade (GEM) in Petri dishes on ice. The released nuclei were collected with a 1 mL cut tip, filtered through a 30 µm nylon net filter (Millipore) and stained with 2 µg/mL DAPI (Merck). Nuclei populations were analyzed using a FACSCanto II High Throughput Sampler cytometer (Becton Dickinson). 10,000 events were measured and each experiment was repeated twice. Data were analyzed using FlowJo v10.1rS software (FlowJo) as follows: FL7-A (DAPI detector) and FSC-A (light scattered forward) were used to gate the different populations. FL8-A and FL8-W (DAPI detectors in a linear scale, detecting area and width) allowed us to discriminate single events from aggregates. Once the singlets were gated, a histogram against FL7-A was plotted and the percentage of nuclei in 2C, 4C, 8C and 16C was measured.

Immunocytochemical and immunohistochemical assays

For immunocytochemical assays (ICC), MM2d cells were collected at 4 days after subculture and fixed in 4% paraformaldehyde in microtubules stabilizing buffer (MTSB; 50 mM PIPES, pH 6.9, 5 mM EGTA, 5 mM $MgSO_4$), for 10 min plus 5 min with vacuum infiltration (30,000 Pa). Cells were washed with MTSB, PBS and water, and then air-dried on Superfrost plus slides (Thermo Scientific). Cells were re-fixed in 4% paraformaldehyde in MTSB for 30 min and washed with MTSB.

For immunohistochemical assys (IHC), either 4 or 7 dps seedlings were fixed in 4% paraformaldehyde in microtubules stabilizing buffer (MTSB; 50 mM PIPES, pH 6.9, 5 mM EGTA, 5 mM $MgSO_4$) for 20 min with vacuum infiltration (30,000 Pa). After several washes with MTSB, PBS and water, seedlings were placed on Superfrost plus slides (Thermo Scientific) and air-dried overnight.

From this point ICC and IHC samples were equivalently treated. Plant cell walls were partially digested with 20 mg/mL driselase (Sigma) in MTSB for 45 min at 37 °C and the slides were washed with PBS. Membranes were permeabilized with 10% DMSO, 3% Igepal CA-630 in MTSB for 1 h. Non-specific sites were blocked in 3% BSA, 10% Horse Serum (HS) in PBS for 1 h at 37 °C. H3K9me2 and H3K27me1 were detected with antibodies (Abcam ab1220 and Millipore 07-448, respectively) diluted 1:1,000 in 1% BSA, 10% HS, 0.1% Tween-20 in PBS

overnight at 4 °C. Slides were washed with 3% BSA in PBS and incubated with donkey anti-mouse 555 and anti-rabbit 488 (A-31570 and A-21206 Life Technologies, respectively) diluted 1:500 in 1% BSA, 10% HS, 0.1% Tween-50 in PBS for 1 h. Following washes in 3% BSA in PBS, nuclei were counterstained with 10 µg/mL DAPI (Merck), washed with PBS and mounted in Mowiol 4-88 (Sigma). The localization of H3K9me2 and H3K27me1 in immunostained cells or roots was analyzed by confocal microscopy (LSM710 Zeiss).

Chasing the cell cycle

To track cell cycle progression either 4 or 7 dps Col-0 wild type, GFP-tagged lines or *ORC1* mutant plants were used. Nuclei undergoing S-phase were labeled with thymidine analogs either EdU (5-ethynyl-2-deoxyuridine; Life Technologies) or BrdU (5-bromo-2-deoxyuridine; Sigma). All the incubations were performed in liquid MS, at room temperature and protecting the samples from light to avoid the degradation of the analogs. The detection of S-phase nuclei was carried out with a single 15 min pulse of 200 µM EdU. For detection of S-phase progression or G2 length, the combination of two pulses was used. First, a 15 min pulse with 200 µM BrdU was done. Then, the analog was washed off and the seedlings incubated with 200 µM thymidine (Sigma) to allow cell cycle progression and avoid new BrdU incorporation. After the chase time, a second pulse with 200 µM EdU was done. Consequently, after the double labeling strategy, cells exclusively marked with EdU were the ones in S-phase, and those stained with BrdU correspond to cells that have progressed during S-phase (BrdU positive, EdU positive) or into the G2 phase (BrdU positive, EdU negative; Fig. 9A-C).

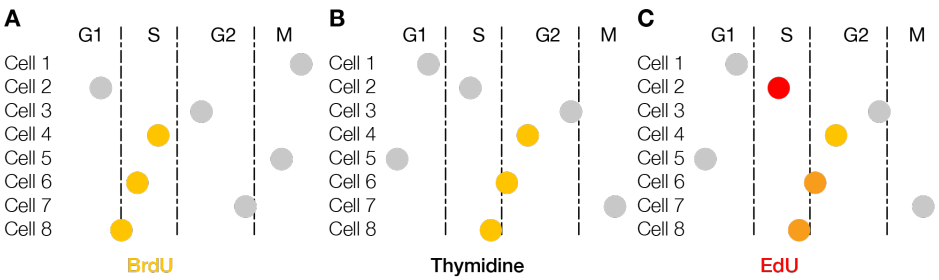


Figure 9 – Cell cycle chase. (A) An asynchronous population of cells incorporates BrdU during S-phase. **(B)** Cell cycle was chased for the desired time in the presence of thymidine to avoid new BrdU incorporation. **(C)** The cells in S-phase incorporate EdU. Four labeled cells types are: not labeled (grey), cells in S-phase (red, only EdU, orange, BrdU and EdU) or G2 cells (yellow, BrdU).

Once roots are labeled, they were fixed according to IHC protocol. After the immunodetection of the GFP in the tagged lines (primary antibody: 1:2,000 anti-GFP, A6455 Life Technologies; secondary antibody: 1:500 goat anti-rabbit 488, A11034 Life Technologies) EdU was detected using Click-iT EdU Alexa Fluor 647 Imaging kit (Life Technologies) following

manufacturer's instructions for 30 min at room temperature. When BrdU is detected, prior to the immunodetection of the proteins and the BrdU itself, DNA was relaxed through a mild digestion by incubating with 0.005 U/µL DNase I RNase free (Roche) for 1h 30 min at 37 °C. DNase I was inactivated with several washes of ice-cold 8 mM EDTA-PBS. BrdU was immunodected using anti-BrdU (347580, Becton Dickinson) diluted 1:200 and donkey anti-mouse-555 (A31570, Life Technologies) diluted 1:500.

3.4 Data analysis

Analysis of confocal microscopy data

Images were processed using ImageJ Fiji v2.0.0-rc/59 and Photoshop CC v2015.5.1 software. Statistical analyses were performed using Prism v5.0a. Figures were made using Illustrator CC v2017.1.0 software.

Positioning of cells along the root

Because the end of the meristem for each cell file is not the same, the meristem size was determined in every cell file of the epidermis. To do so, the distance from the QC to the first cell in focus in the epidermis was measured as well as the length of all the other cells in the file. The end of the meristem was considered as the first elongated cell (Casamitjana-Martinez et al., 2003). The relative position of each cell was calculated by normalization to the length of the RAM in each cell file, thus, in this analysis, the cell marking the end of the meristem will be at position 1 and 0 corresponds to the QC. The relative position was calculated according to:

$$\text{Relative position} = \frac{(\text{Cell length} + \text{Initial distance from the QC})}{(\text{Length of cell determining the end of the meristem} + \text{Initial distance from the QC})}$$

Epidermal T-clones and cells undergoing mitosis were removed from the analysis.

Measurements of fluorescent intensity

The fluorescent intensity was measured as the integrity density of a determined ROI or along a line as the Gray Value. In all the cases independent measurements were taken for each color channel.

BrdU sequencing data analysis

Genomic DNA was purified from BrdU-labeled cells over CsCl gradients. DNA in the heavy-light and in the light-light fractions was used as sample and control, respectively. BrdU sequencing data reads (GEO GSE2182; (Costas et al., 2011)) were trimmed down to 50 nt from the 3' end and mapped to the reference Arabidopsis genome (TAIR10) using BOWTIE aligner (Langmead et al., 2009), allowing up to three mismatches and discarding multihit reads. PCR duplicate reads were removed using an in-house script. Peak calling was performed using six different peak calling algorithms, namely, MACS (Zhang et al., 2008); versions 1.4 and 2.0), BayesPeak (Spyrou et al., 2009), T-PIC (Hower et al., 2011), HOMER (Heinz et al., 2010) and SICER (Zang et al., 2009). Neighboring peaks were merged when interpeak distance was less than 260 nt. Peaks smaller than 200 nt were removed from the analysis. The same analysis was carried out using only the multihit reads using MACS1.4.

ORI distribution and classification

General annotation coverage was calculated with the complete set of annotations from TAIR10, discarding "transposon_fragment" as it is redundant with the "transposable_element" annotation. Pericentromeric regions were defined as the regions where the gene coverage in 1 Mb bin was equal or lower to 40%. ORIs were attributed to a type of annotation (genes, TEs or particular TE families) only for unambiguous non-overlapping annotation. TE family coverage was calculated within the TE genome space (total TE nucleotide content).

C methylation, G quadruplex, GC content and chromatin states analysis

CG, CHG and CHH methylation data were retrieved from (GEO GSE39901) (Stroud et al., 2013). The presence of G quadruplexes in the Arabidopsis genome was predicted using the Quadparser software (Hershman et al., 2008) allowing a spacing of 7 nt between G-strings. The GC content of the genome was calculated in bins of 50 nt. For the analysis of the distribution of TE among the different chromatin states (Sequeira-Mendes et al., 2014), the relative frequency of each TE family in each state was determined by the coverage of the family in that particular state relative to the total coverage of the TE family in the genome. For the distributions of ORI-TEs among the different chromatin states the ORI midpoint was considered. All the bioinformatics analyses were performed with in-house Perl scripts and BEDtools suite utilities (makewindows, genomecov, merge, intersectBed) (Quinlan, 2014).

4. Results

4.1. The two Arabidopsis *ORC1* genes are differentially regulated

In silico characterization of the upstream regions of the *ORC1* genes

In Arabidopsis, two genes encode the largest subunit of the *ORIGIN RECOGNITION COMPLEX* (*ORC*; (Diaz-Trivino et al., 2005)). The two *ORC1* genes are quite similar both in DNA (90% identity) and protein sequences (88% identity). However, the genomic regions upstream the genes are utterly different. We used a genome browser to analyze different aspects of the chromatin that would help us define the promoter region for each gene. First, we looked at the features defined in the genome. Interestingly, the upstream region of *ORC1a* (AT4G14700) contains a transposable element (TE; AT4TE37370) just 87 bp upstream of the gene (Fig. 10A).

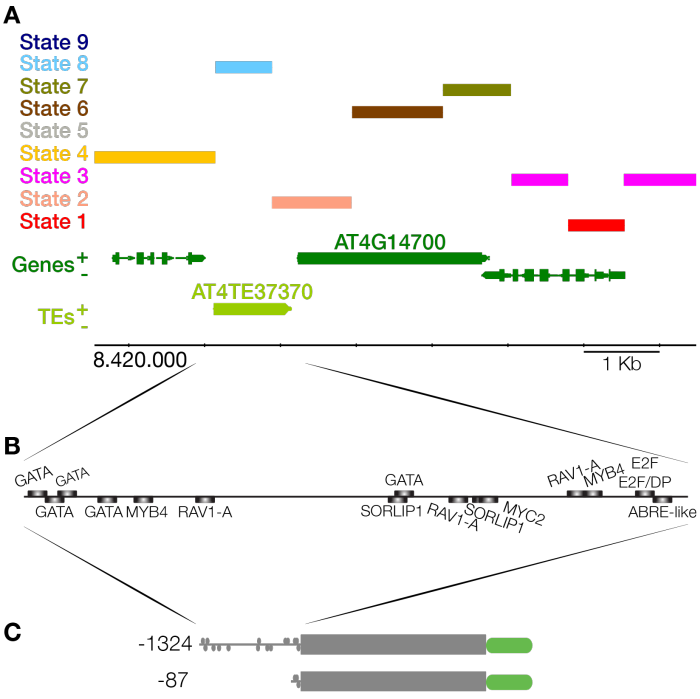


Figure 10 – *ORC1a* genomic landscape. (A) Genome-browser view of the region surrounding *ORC1a* locus at the chromosome 4. Genes (dark green) transcribed from each strand, TEs (light green) and the nine chromatin states are shown along the chromosome together with the coordinate scale. Chromatin states, largely corresponding to various genomic elements, are as follows: state 1 (red), TSS; state 2 (salmon), proximal promoters; state 3 (pink), 5' half of genes; state 4 (yellow), distal promoters enriched in H3K27me3; state 5 (grey), Polycomb-regions; state 6 (brown), average gene bodies; state 7 (olive green), long gene bodies; state 8 (light blue), AT-rich heterochromatin; state 9 (dark blue), GC-rich heterochromatin. **(B)** Close view of the upstream region of *ORC1a* showing the predicted TF binding sites. **(C)** Representation of the cloned constructs harboring the promoters and gene (gray) and the tag (green) for ORC1a analysis.

It has been previously described that TEs that are close to a gene can influence its expression (Lippman et al., 2004); (Hirsch and Springer, 2017). Despite the active chromatin state 2 of the proximal promoter and TSS of *ORC1a*, the upstream region is in a closed, AT-rich heterochromatic conformation due to the TE (chromatin state 8). The region between the TE and the gene upstream *ORC1a* possesses the chromatin state 4, corresponding to distal regulatory elements that can be affecting the expression of *ORC1a* (Sequeira-Mendes et al., 2014); Fig. 10A). We then searched *in silico* the transcription factor (TF) binding sites for this

particular genomic area (Hudson and Quail, 2003) and found three main regions enriched in TFs, before, inside, and after the TE. Just upstream *ORC1a* there are two E2F binding sites (position -68) and then others related to plant hormones (ABRE-like), growth and developmental factors (GATA, MYB4, RAV1-A; Fig. 10B). We decided to clone 1324 bp upstream of *ORC1a* as the promoter region (Fig. 10C). However, the construct harboring this long promoter was found active in only 7% of the lines analyzed (from a total of 42 lines). Subsequently we cloned a short promoter version, which lacked the TE and consisted of only 87 bp in length (Fig. 10C), and found that 67% of the lines containing the short promoter were active (from a total of 18 lines). This revealed that the presence of the TE in the transgenic construct significantly affected *ORC1a* expression.

In contrast, the chromatin upstream *ORC1b* (AT4G12620) is in a more open conformation (Fig. 11A), and chromatin states 1 and 2, typical of proximal promoters and TSS, can be found between *ORC1b* and the close adjacent gene. When we studied the putative TF binding sites (Fig. 11B) we found two E2F sites very close to the TSS (position -104). Curiously, *ORC1b* contains many TF sites related to light (SORLIP), growth and developmental factors (GATA, MYB4). We decided to clone the region between *ORC1b* and the neighbor gene, resulting in a 793 bp promoter length (Fig. 11C).

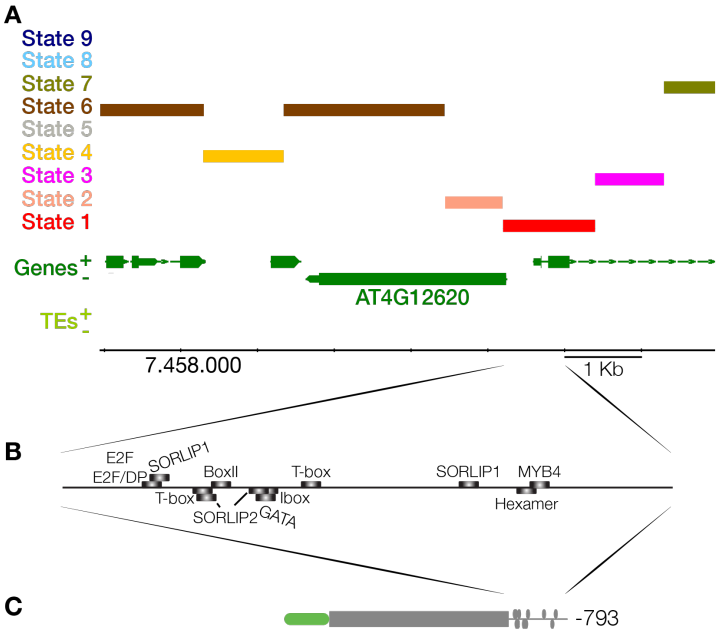


Figure 11 – *ORC1b* genomic landscape. (A) Genome-browser view of the region surrounding *ORC1b* locus at the chromosome 4. Genes (dark green) transcribed from each strand, TEs (light green) and the nine chromatin states are shown along the chromosome together with the coordinate scale. Chromatin states, largely corresponding to various genomic elements, are as follows: state 1 (red), TSS; state 2 (salmon), proximal promoters; state 3 (pink), 5' half of genes; state 4 (yellow), distal promoters enriched in H3K27me3; state 5 (grey), Polycomb-regions; state 6 (brown), average gene bodies; state 7 (olive green), long gene bodies; state 8 (light blue), AT-rich heterochromatin; state 9 (dark blue), GC-rich heterochromatin. **(B)** Close view of the upstream region of *ORC1b* showing the predicted TF binding sites. **(C)** Representation of the cloned construct harboring the promoter and gene (gray) and the tag (green) for ORC1b analysis.

Expression pattern domains of ORC1 proteins

In previous studies based on gene expression analysis at the mRNA level, it was found that, while *ORC1a* was expressed at the endocycling cells, *ORC1b* expression was restricted to the proliferating ones, without having ruled out the possibility that *ORC1b* could be also present in endocycling cells (Diaz-Trivino et al., 2005). We wanted to study the regulation at the protein level and therefore created different translational reporter lines. To define the expression domains of both ORC1 proteins we first analyzed the pattern of the GUS fusion constructs. In young seedlings, ORC1a-GUS was present just above the root apical meristem (RAM) at low levels and specially concentrated in endocycling cells in the transition zone (Fig. 12A). On the contrary, ORC1b-GUS was also present in the proliferating cells with a higher intensity than in the endocycling cells (Fig. 12B). In the shoot apical meristem (SAM), ORC1a-GUS is detected at early time points (4 dps) only in a few cells, perhaps undergoing the endocycle in growing leaves (Fig. 12C, D). As for ORC1b-GUS, that is present in the SAM, it is visible since early development, including cells of the stomatal lineage in cotyledons (Fig. 12E, F). Once the aerial part starts to develop (12 dps), ORC1a-GUS is detectable only at the cells undergoing endocycle (Fig. 12G, H), while ORC1b-GUS is present also in the proliferating cells as those of the basal half of the leaves (Fig. 12I, J). At this stage, ORC1b-GUS is more present in the endocycling cells of the leaves and trichomes than ORC1a-GUS. The observed protein expression domains are in agreement with previous data of RNA levels (Diaz-Trivino et al., 2005).

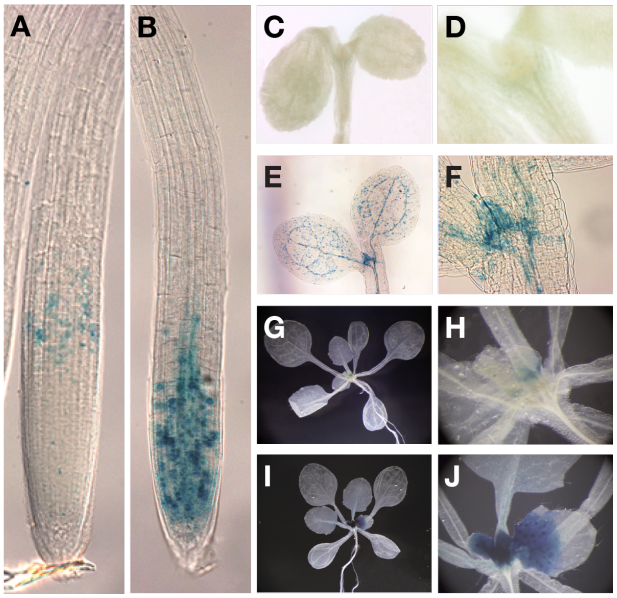


Figure 12 – Expression domains of the ORC1 proteins in young seedlings. Detection of ORC1a-GUS (*pORC1a(-1324)::ORC1a-GUS*) at **(A)** 7 dps root apical meristem (RAM), **(C, D)** 4 dps shoot apical meristem (SAM) and **(G, H)** 12 dps SAM. Presence of ORC1b-GUS at **(B)** 7 dps root apical meristem (RAM), **(E, F)** 4 dps shoot apical meristem (SAM) and **I, J)** 12 dps SAM.

To determine ORC1 location at the subcellular level, we used translational ORC1-GFP fusion constructs, in which *ORC1* genes were expressed under their own promoters. Analysis of

these reporter lines by confocal microscopy of the RAM confirmed that the two ORC1 proteins exhibit distinct expression domains in nucleus of root cells: ORC1a was specifically detected in the endocycling and first elongating cells prior to exit to differentiation (Fig. 13A, B) whereas ORC1b was detected in proliferating cells as well as in those of endocycling cells in the transition zone and the first elongating cells (Fig. 13B). Remarkably, the expression pattern of ORC1a did not change between the short (-87; Fig. 13A) and long promoter (-1324; Fig. 13B) versions. The simplified diagram in Fig. 13D shows the three different zones of the tip of the root and where cell cycle and endocycle programs take place.

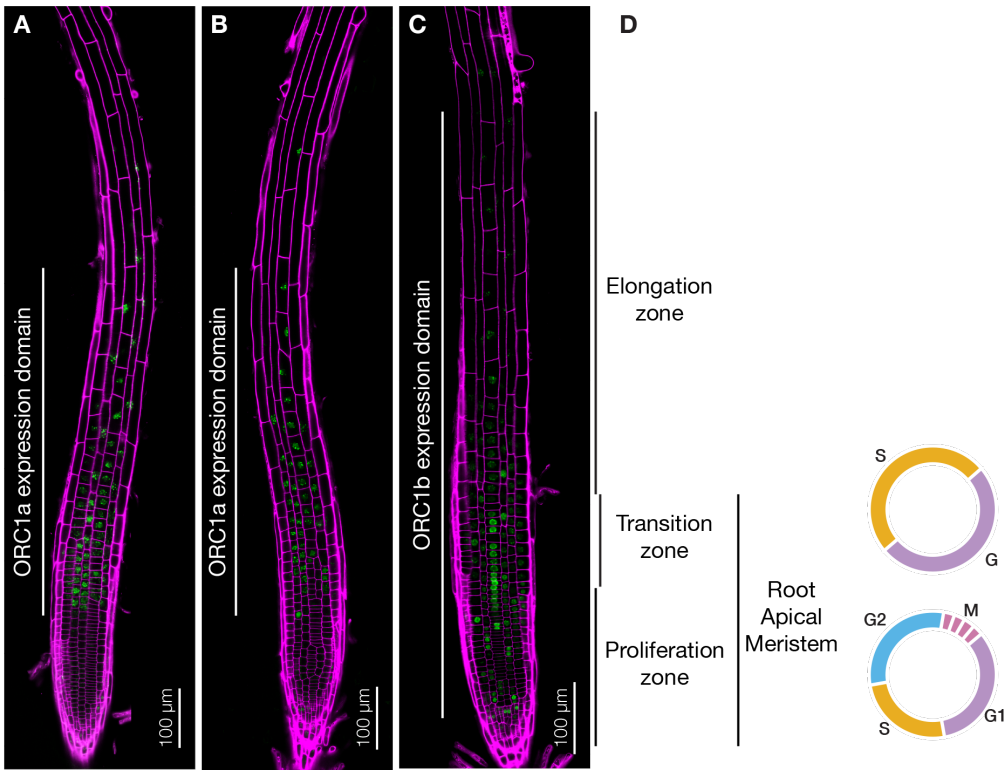


Figure 13 – Expression domains of ORC1 proteins at the RAM. The cell wall of 7 dps roots was stained with propidium iodide (magenta) and GFP was directly detected under the confocal microscope: **(A)** *pORC1a(-87)::ORC1a:GFP* expression is detected at the transition to elongation zone. **(B)** *pORC1a(-1324)::ORC1a:GFP* displays the same expression domain than the long promoter version. **(C)** *pORC1b::ORC1b:GFP* exhibits expression in both the proliferation, transition and first part of the elongation zones. **(D)** Sketches disclose the zones at the root tip and the distribution of cell cycle and endocycle, as well as their phases marked in purple (G1 or G), orange (S-phase), blue (G2) and pink (mitosis).

We also examined the quiescent center (QC) and the niche of stem cells in the RAM and found them completely depleted of ORC1a (Fig. 14A). Interestingly, only a small proportion of cells contained ORC1b (Fig. 14B, C), likely reflecting their low proliferation status.

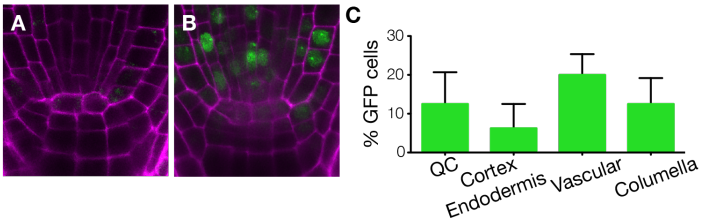


Figure 14 – GFP expression at the RAM stem cell niche. (A) Absence of ORC1a-GFP signal. **(B)** ORC1b-GFP expression in root stem cells. **(C)** Percentage of positive ORC1b-GFP cells in the QC (quiescent center), cortex-endodermis, vascular and columella initial cells. The mean and the standard error of the mean is plotted, n=8 roots.

A closer inspection of the GFP labeled nuclei revealed the occurrence of different patterns for ORC1a as well as for ORC1b, strongly suggesting that they are regulated during the cell cycle and the endocycle. In order to quantify the distinct patterns, we measured epidermal cells along the same cell file from the QC to the last cell at the elongation zone prior to differentiation. Because the end of the meristem for each cell file is not the same, the distance from the QC was normalized to the length of the RAM in each cell file, as the point where cell length doubles (Casamitjana-Martinez et al., 2003). Thus, 0 corresponds to the QC and 1 to the end of the meristem (Fig. 15A, B). We found that most of the cells in the RAM were depleted of ORC1a-GFP (92.7%, Fig. 15A, C), and those labeled with GFP signal have three different patterns, colocalizing with the entire nucleus (euchromatic, 3.8%), clearly at the chromocenters (heterochromatic, 1.7%) or at the chromocenters plus a slight signal at the euchromatin (both, 1.8%). This distribution changed dramatically when we looked outside of the meristematic zone (Fig. 15A, D). At the elongation region only 36.5% of the cells appeared empty of ORC1a and the majority of the GFP positive cells contain the signal at the chromocenters (42.1%). Just in a few cells ORC1a-GFP could be detected only at the euchromatin (1.6%) being more frequent to present the protein at both chromocenters and euchromatic sites (19.8%). Nevertheless, the distribution of the cells marked with GFP only at the euchromatin was restricted to the surroundings of the end of the meristem (Fig. 15A). These results led us to conclude that ORC1a is primarily expressed in the endocycle zone of the root and largely associated with chromocenters that contain constitutive heterochromatin.

Regarding ORC1b we found that inside the RAM (Fig. 15B, E), half of the nuclei contained a clear GFP signal at the chromocenters with a more diffuse labeling of the rest of the nucleus (both, 28.2% with a low intensity and 21.9% with a high intensity) whereas almost the other half were totally depleted of ORC1b-GFP (empty, 40.7%). In addition, some nuclei showed the protein homogeneously labeling the entire nucleus (euchromatin, 5.6% with a low intensity and 1.3% with a high intensity), while just a few were only labeled at the chromocenters (2.3%). In

the upper part of the root (Fig. 15B, F) the number of cells depleted of ORC1b decreased (21.2%), while those with the signal present at both chromocenters and euchromatin (29.7% low euchromatin or 21.8% high euchromatin labeling) is maintained. The most notorious change was for the nuclei containing ORC1b at the chromocenters, which accounted for 26.0% of the total nuclei analyzed above the meristem. There was also a significant reduction of the nuclei with labeling of ORC1b at the euchromatin (1.0% with a low intensity and 0.3% with a high intensity). Like in the case of cells marked with ORC1a exclusively at the euchromatin, the cells with only ORC1b euchromatic labeling were restricted to the RAM and the cells just above the meristem.

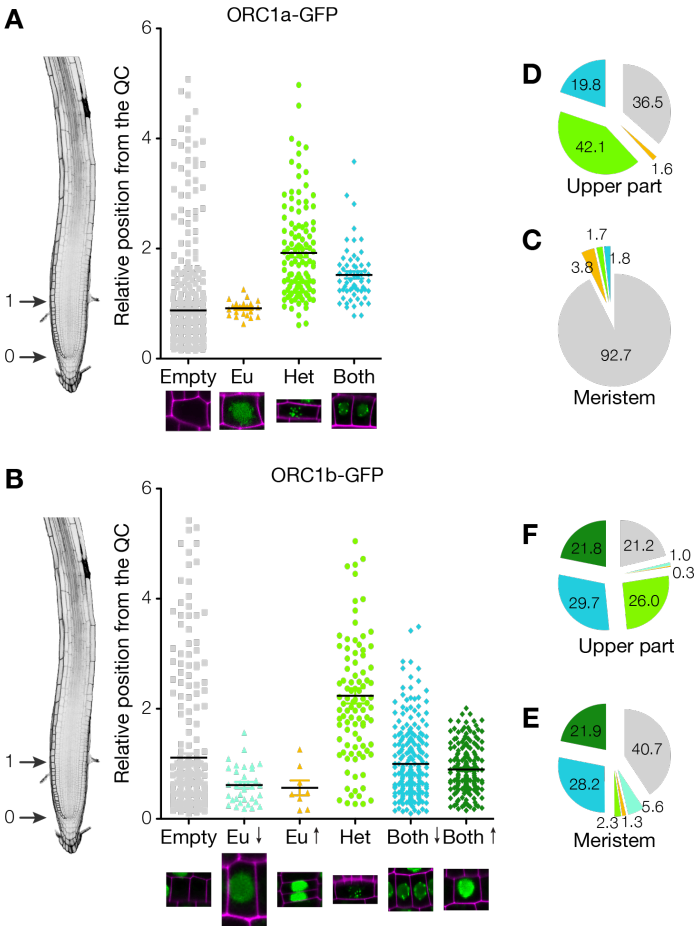


Figure 15 – Pattern distribution of ORC1a and ORC1b along the root. (A) ORC1a-GFP signal distribution in four main patterns, namely empty (grey), Eu (euchromatin, orange), Het (chromocenters, green) and Both (chromocenters and euchromatin, blue). **(B)** ORC1b-GFP signal distribution along the root in six patterns, namely empty (grey), Eu ↓ (euchromatin low signal, light blue), Eu ↑ (euchromatin high signal, orange), Het (chromocenters, light green), Both ↓ (chromocenters and euchromatin low signal, dark blue) and Both ↑ (chromocenters and euchromatin high signal, dark green). At the left of each graph, a root shows the end of the meristem measured at the epidermal layer. Nuclei patterns along the root were normalized to the end of meristem (1) relative to the distance from the QC (0). Under the graph there are representative examples of each pattern. Pie charts show the percentage distribution for each pattern of ORC1a-GFP at the meristem **(C)**, the upper part of the root **(D)**, or the patterns of ORC1b-GFP at the meristem **(E)** and the upper part of the root **(F)**. For each construct five roots of 7 dps plants were analyzed. The mean (black bar) and the standard error of the mean are plotted.

These localization studies reveal that (i) both ORC1 proteins are nuclear and have different expression domains, ORC1a being specifically located in endocycling cells and ORC1b in both proliferating and endocycling cells, (ii) the different nuclear accumulation of ORC1a or ORC1b may reflect their status during the progression of the cell cycle and the endocycle, and (iii) the overall distinct expression strongly suggest that ORC1a and ORC1b may play different roles during proliferation and endocycle.

ORC1b starts to accumulate during G2

The different nuclear patterns of ORC1b in proliferating cells strongly suggested us that it is highly regulated during cell cycle progression. To determine this and define precisely its timing we used live imaging of ORC1b-GFP roots. Since we noticed several mitotic figures in the roots labeled with ORC1b-GFP, we first studied the kinetics during mitosis. Root membranes were stained with FM4-64 and images were taken every two min for at least 30 min to follow the entire process (Fig. 16A, B).

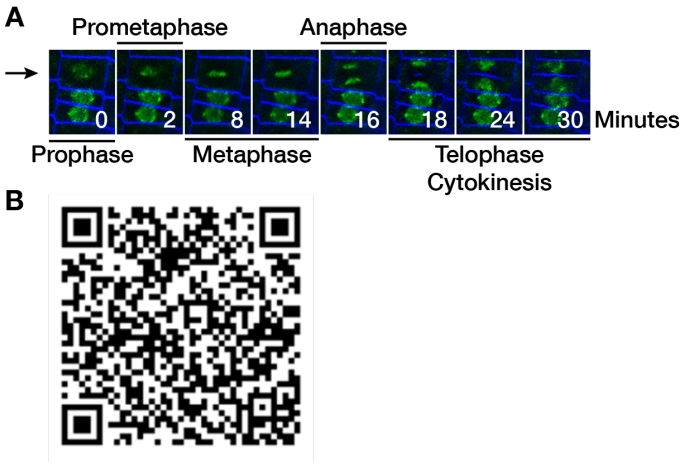


Figure 16 – Live imaging of cells during mitosis. ORC1b-GFP roots were stained with FM4-64 and consecutive pictures were taken with the confocal microscope, every two min for a total of 30 min. **(A)** Representative pictures of all the phases of the mitotic process. **(B)** QR code to access the complete video.

We found that, ORC1b-GFP was not only present at the end of the mitosis but since the beginning in prophase (Fig. 16A, B). Based on the time covered by this experiment it is conceivable that ORC1b would be loaded at some point in G2 cells, as recently shown for human Orc1 (Kara et al., 2015). To confirm this we used a double labeling strategy to unequivocally identify G2 cells in the growing root. Seedlings were labeled first with a 15 min pulse of BrdU, then BrdU was washed off and cell cycle progressed 2h in the presence of thymidine to block new BrdU incorporation. Finally, a second 15 min pulse with EdU marked the cells in S-phase. This allowed us to distinguish unambiguously cells in G2 as those uniquely labeled with the first analog, BrdU. Using this strategy we found that some G2 cells (BrdU+, EdU-) contain ORC1b-GFP (Fig. 17).

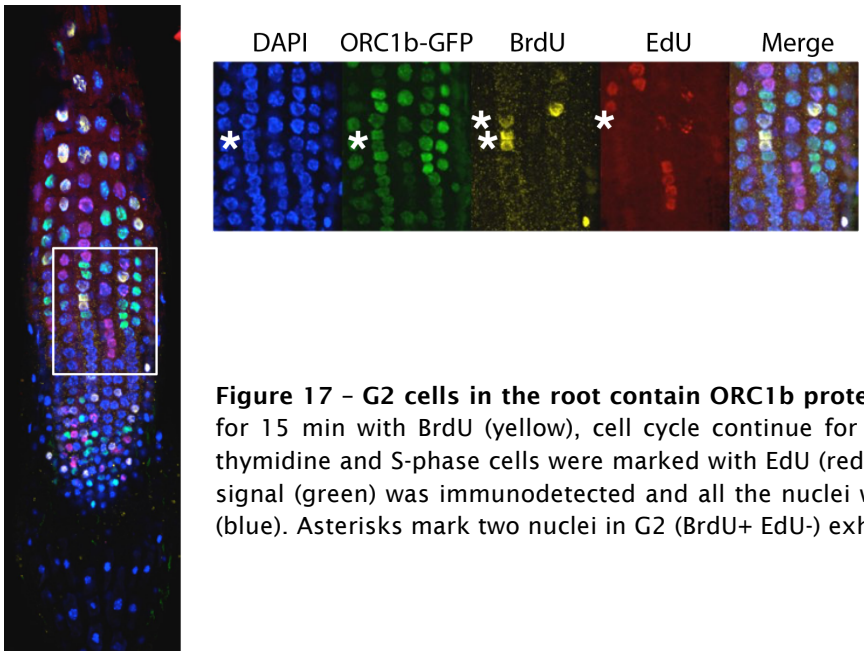


Figure 17 – G2 cells in the root contain ORC1b protein. Roots were labeled for 15 min with BrdU (yellow), cell cycle continue for 2h in the presence of thymidine and S-phase cells were marked with EdU (red) for 15 more min. GFP signal (green) was immunodetected and all the nuclei were stained with DAPI (blue). Asterisks mark two nuclei in G2 (BrdU+ EdU-) exhibiting ORC1b-GFP.

Cyclins are proteins expressed at specific time points of the cell cycle (Gutierrez, 2009). The transcriptional fusion construct of the promoter of *CYCB1;1* with GFP makes possible to detect cells in G2 phase at the root (Colon-Carmona et al., 1999; Ubeda-Tomas et al., 2009). We crossed our plants expressing ORC1b fused to mRFP with the *pCYCB1;1::GFP* reporter line and analyzed the roots under the confocal microscope. We noticed that the cells in G2 (CYCB1;1-GFP) were positive for ORC1b-mRFP with low intensity (Fig. 18), corresponding to the patterns observed euchromatin low intensity and chromocenters and euchromatin low intensity (Fig. 15B).

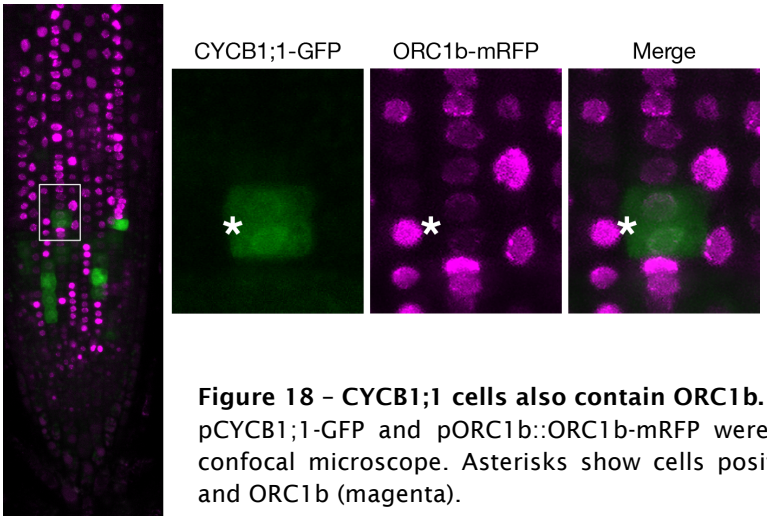


Figure 18 – CYCB1;1 cells also contain ORC1b. Plants with the two constructs pCYCB1;1-GFP and pORC1b::ORC1b-mRFP were directly observed under the confocal microscope. Asterisks show cells positive for both *CYCB1;1* (green) and ORC1b (magenta).

Together, this set of data demonstrates that (i) some proliferating cells do not contain any detectable ORC1b at some point during the cell cycle, (ii) ORC1b starts to be loaded in G2, and (iii) remains bound to the chromatin for the entire mitosis.

Full loading of ORC1b occurs in early G1

To further assess the dynamics of ORC1b protein during G1, we focused on cells that have just completed cytokinesis. Live imaging experiments (Fig. 19A, C) revealed that the intensity of the ORC1b-GFP signal increases right after cells exit mitosis, reaching a plateau level approximately one hour afterwards (Fig. 19B).

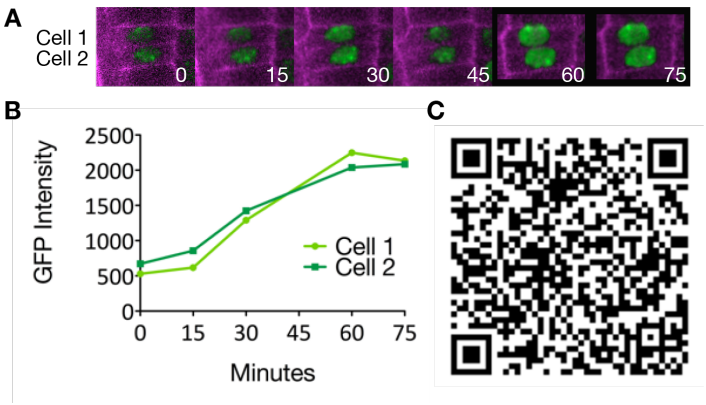


Figure 19 – Live imaging analysis of cells after cytokinesis. Roots were stained with propidium iodide (magenta) and ORC1b-GFP was directly detected, collecting pictures every 15 min. **(A)** ORC1b-GFP from cells one and two increases after mitosis. **(B)** The GFP intensity reaches a plateau one hour after the exit of mitosis. **(C)** QR code to access the complete video.

This result demonstrates that, although ORC1b is first synthesized during G2, the major amount of the protein is loaded onto the chromatin during G1 phase, coinciding with ORI licensing.

ORC1b is rapidly degraded shortly after G1/S transition

To follow ORC1b dynamics later in the cell cycle, we determined the ORC1b status in cells undergoing S-phase by labeling with EdU for 15 min. We found that the vast majority of cells that were in S-phase (EdU+) did not contain any detectable ORC1b-GFP signal (only ~3% of EdU+ cells were also GFP+, Fig. 20).

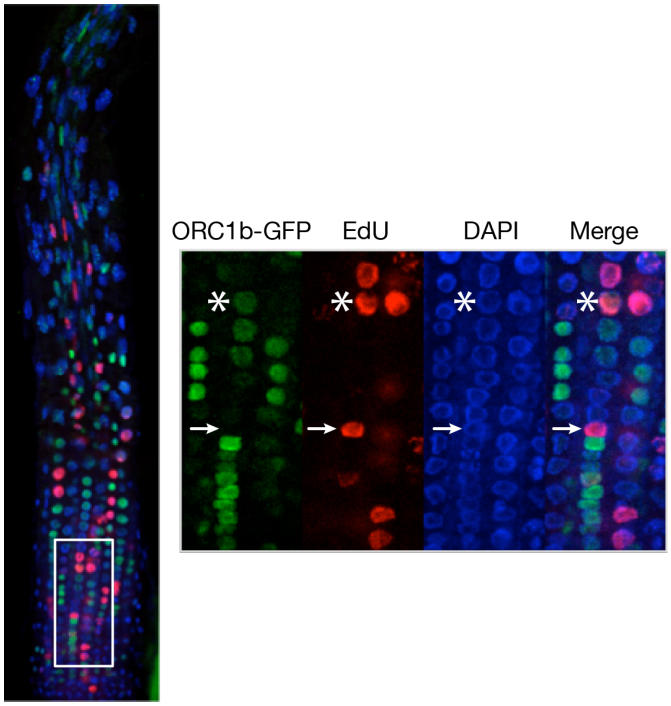


Figure 20 – S-phase cells in the root are depleted in ORC1b protein. Roots were labeled for 15 min with EdU (red), GFP signal (green) was immunodetected and all the nuclei were stained with DAPI (blue). Asterisks mark cells entering S-phase labeled with ORC1b-GFP and EdU. Arrows indicate cells in S-phase, positive for EdU but depleted of ORC1b-GFP.

This indicates that ORC1b is not present during most of the S-phase and suggests that an active degradation of ORC1 occurs shortly after the G1/S transition. Live imaging experiments demonstrated that this is likely the case (Fig. 21A, C) because nuclei displaying high intensity ORC1b-GFP signal, fully lost it in less than ten min (Fig. 21B).

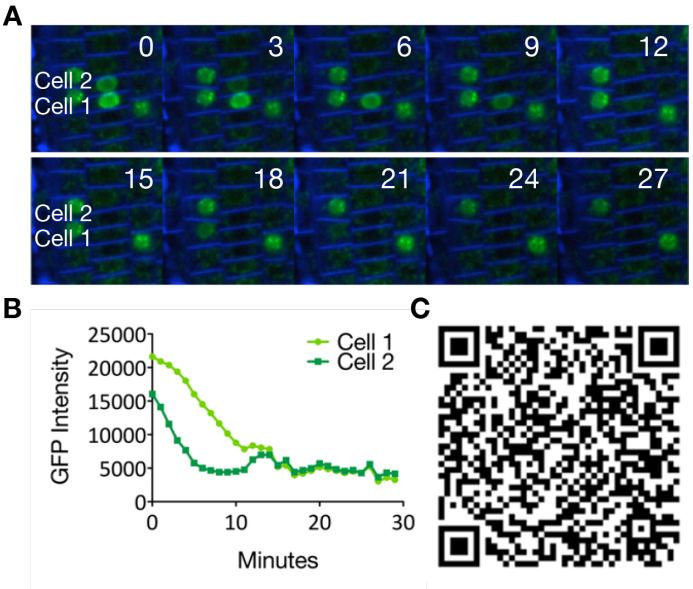


Figure 21 – Degradation of ORC1b-GFP. ORC1b-GFP roots were stained with FM4-64 and consecutive pictures were taken with the confocal microscope, every minute for a total of 30 min. **(A)** ORC1b-GFP from cells one and two rapidly disappear. **(B)** The GFP intensity is completely lost in less than ten min. **(C)** QR code to access the complete video.

These experiments indicate that (i) cells in G1, with high amounts of ORC1b, entirely lose it at the transition to S-phase, and (ii) during the S-phase cells are depleted of ORC1b.

ORC1b, but not ORC1a, is degraded by a SCF^{FBL17} E3 ligase

Cell cycle regulated proteins are often degraded through the ubiquitin-proteasome pathway, as it is the case, for example, for Orc1 in human cells (Mendez et al., 2002). To evaluate whether ORC1 proteins are targeted for proteasome degradation we first treated the GFP-tagged expressing plants with various proteasome inhibitors. Roots in the presence of MG132 and epoxomicin, that targets the proteasome for reversible and irreversible inhibition, respectively revealed that the amount of ORC1a did not change significantly. However, we observed that its subcellular localization pattern was altered, and now ORC1a was mostly located at the euchromatin (Fig. 22A). On the contrary, ORC1b accumulated to the higher levels (Fig. 22B) and the association to the chromocenters was lost.

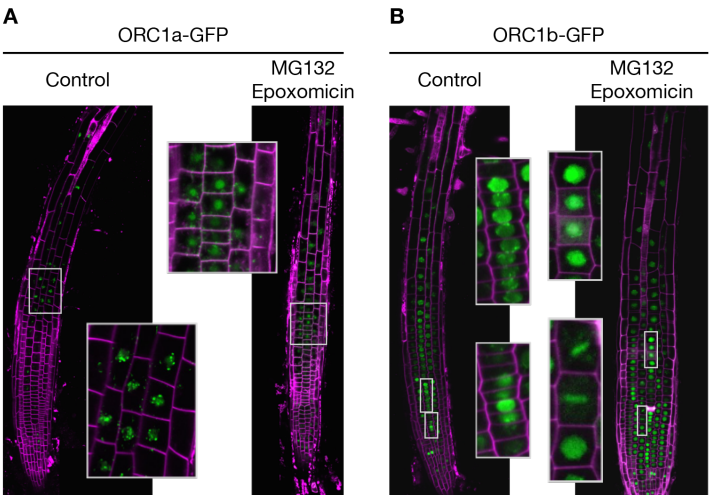


Figure 22 – ORC1b is accumulated after proteasome inhibitor treatment. Plants were treated for four hours with 100 µM MG132 and 1 µM epoxomicin, cell walls were stained with propidium iodide (magenta) and then analyzed at the confocal microscope. **(A)** No changes were detected between the control roots and the treated ones of ORC1a-GFP expressing plants. **(B)** Roots of plants expressing ORC1b-GFP accumulate high amounts of the protein and the typical accumulation at chromocenters is loss.

Inhibiting the E1 NEDD8 activating enzyme with MLN4924 (Hakenjos et al., 2011) essentially had no effect on the overall amount of ORC1a (Fig. 23A), although ORC1b protein was stabilized at its canonical locations, colocalizing with both the euchromatin and the heterochromatin in most of the cells (Fig. 23B).

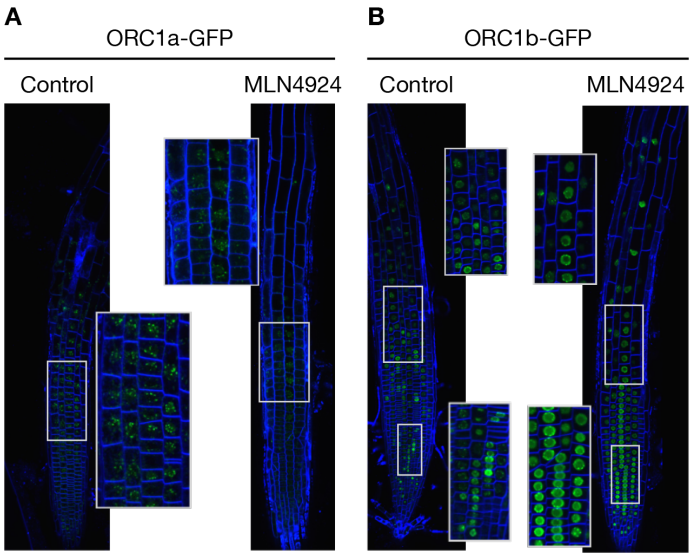


Figure 23 – ORC1b is accumulated after E1 activating enzyme inhibitor treatment. Plants were treated for six hours with 25 µM MLN4924, membranes were stained with FM4-64 (blue) and then analyzed at the confocal microscope. **(A)** No changes were detected between the control roots and the treated ones of ORC1a-GFP expressing plants. **(B)** ORC1b-GFP proteins are stabilized and accumulated high amounts at the roots.

These data clearly demonstrate that ORC1b, but not ORC1a, is targeted for polyubiquitination and proteasome degradation. The change in localization of ORC1a pattern points to a monoubiquitination that would triggers the change in subnuclear distribution.

Several key cell cycle regulatory proteins in animal cell, including E2F1, Orc1, CycE and Cdt1, among others, are targeted by the SCF^{Skp2} E3 ubiquitin ligase for proteasome degradation (Mendez et al., 2002). In Arabidopsis there are two homologs for *SKP2*, *SKP2A* and *SKP2B*, but only *SKP2A* is expressed at the root meristem and was found to regulate the stability of E2Fc (del Pozo et al., 2002) and DPb (del Pozo et al., 2006). Thus, we first tested whether a SCF^{SKP2A}

E3 ligase was relevant for ORC1 stability by expressing ORC1a-GFP and ORC1b-GFP in a *skp2a-1* mutant background. We found that the expression domains and the amount of protein were unmodified for both ORC1a (Fig. 24A) and ORC1b (Fig. 24B), indicating that in Arabidopsis SKP2A does not recognize ORC1 Arabidopsis proteins.

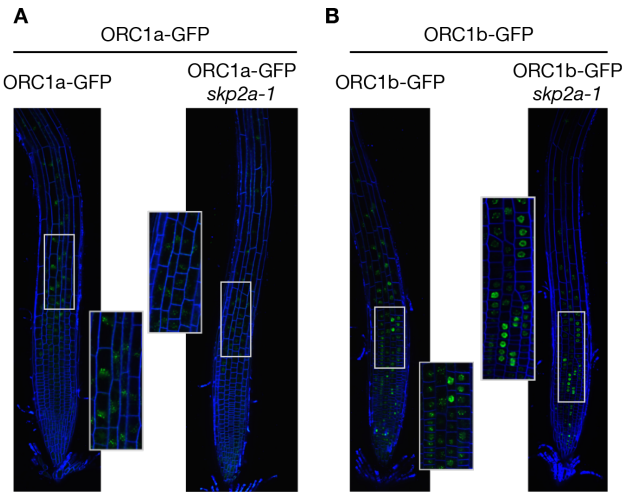


Figure 24 – ORC1-GFP expression remains unchanged in the absence of SKP2A. (A) Plants expressing ORC1a-GFP in wild type and *skp2a-1* mutant background. **(B)** Plants expressing ORC1b-GFP in wild type and *skp2a-1* mutant background. Root membranes were stained with FM4-64 (blue).

Next we sought to identify the F-box protein relevant for the control of ORC1b availability. In plants there are ~600 F-box proteins described (Gagne et al., 2002). It has been recently shown that FBL17 is a key F-box protein controlling the G1/S and G2/M transitions (Noir et al., 2015) as well as germline proliferation (Gusti et al., 2009), being a good candidate for recognition to degradation of ORC1b. In collaboration with the group of Pascal Genschik we crossed the GFP tagged lines with the *fbl17-1* mutant. Plants expressing ORC1a-GFP in a *fbl17-1* mutant background (Fig. 25A) present the same amount of protein, while the ones expressing ORC1b-GFP in a *fbl17-1* mutant background clearly showed an accumulation of the protein in most nuclei (Fig. 25B). Furthermore, the high levels of the protein in the nuclei led to its accumulation at the nucleoli.

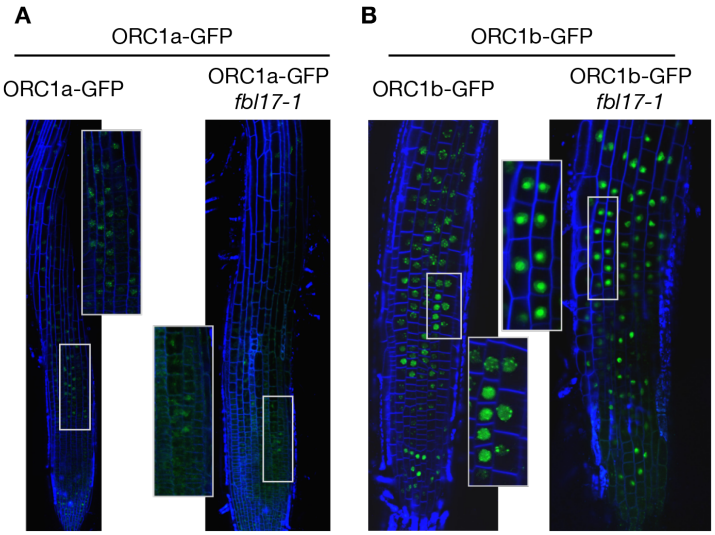


Figure 25 – ORC1b is degraded by the SCF^{FBL17} E3 ligase. (A) Plants expressing ORC1a-GFP in wild type and *fbl17-1* mutant background. **(B)** Plants expressing ORC1b-GFP in wild type and *fbl17-1* mutant background. Root membranes were stained with FM4-64 (blue).

To further determine the relation between FBL17 and ORC1b, live imaging of plants co-expressing FBL17-GFP and ORC1b-mRFP supported a role of FBL17 in regulating ORC1b availability (Fig. 26A, C). Thus, ORC1b levels rapidly decreased coinciding with the accumulation of FBL17 at the G1/S transition (Fig. 26B). Collectively, these data show that ORC1b, but not ORC1a, is targeted for proteasome degradation by an SCF^{FBL17} complex.

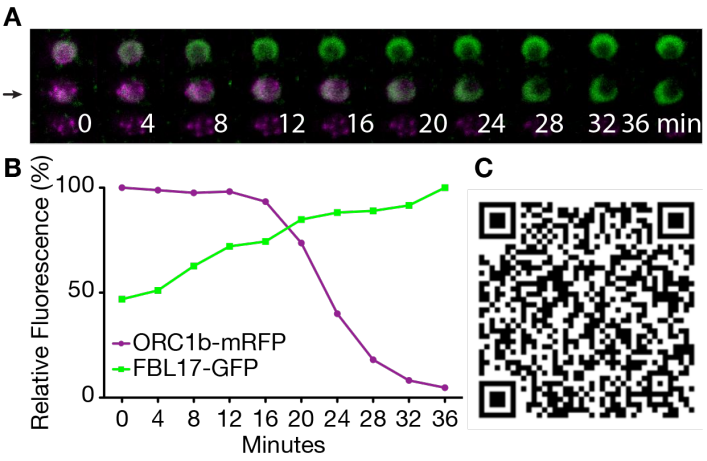


Figure 26 – FBL17 degrades ORC1b (A) Plants expressing both FBL17-GFP (green) and ORC1b-mRFP (magenta) were analyzed by confocal microscopy every four min. **(B)** Relative fluorescent levels of the two tagged proteins shows that a previous accumulation of FBL17 is needed for ORC1b degradation. **(C)** QR code to access the complete video.

Cell cycle regulated proteins are often phosphorylated by CDKs. This is the case of human Orc1 or Cdt1, which are phosphorylated to be marked for polyubiquitination and proteasome degradation. To elucidate if that was the signal that triggers ORC1b degradation at the G1/S transition we treated plants expressing ORC1a-GFP and ORC1b-GFP, using CDT1a-GFP as positive control (Lopez et al., unpublished, (Castellano et al., 2004); Fig. 27) with roscovitin. We found that at early time points when CDT1a started to accumulate, the amount of ORC1a and ORC1b was slightly decreased. At longer time points the ORC1 proteins were barely detectable (Fig.27). This result suggests that the signal for ORC1 degradation is not likely a phosphorylation event dependent on CDK activity.

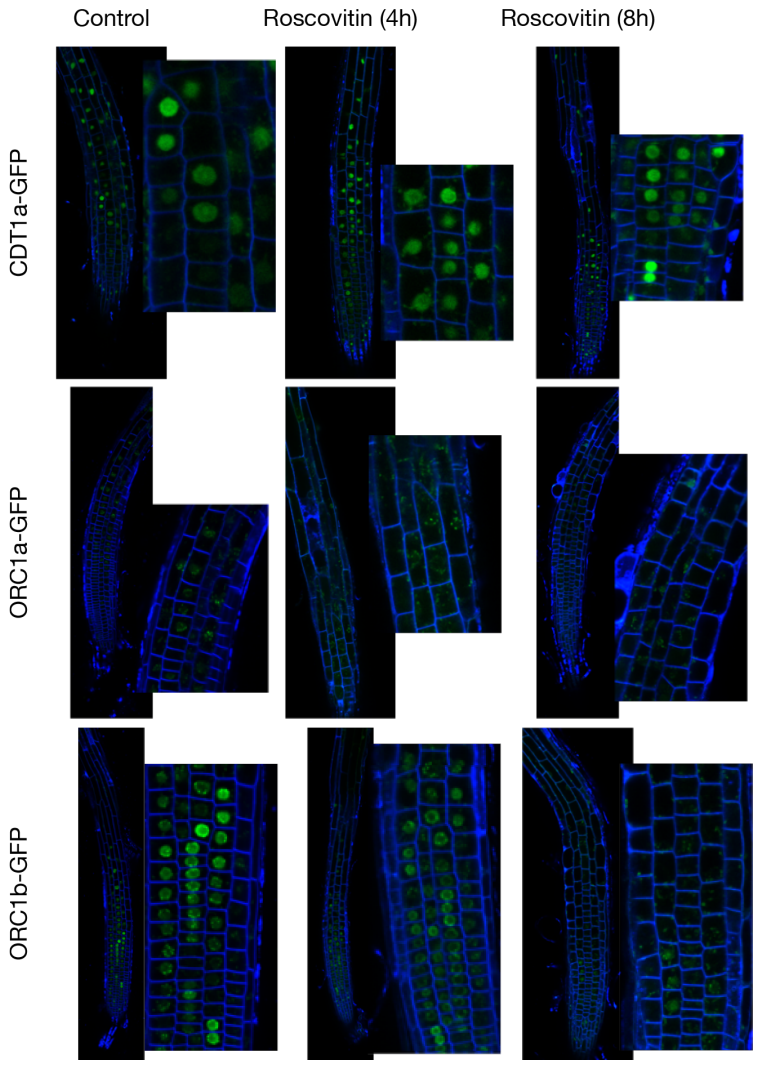


Figure 27 – Under roscovitin treatment ORC1 proteins are degraded. Plants expressing ORC1a-GFP, ORC1b-GFP or CDT1a-GFP were incubated for four or eight hours with 100 µM roscovitin.

Both ORC1 proteins are present during the G-phase of the endocycle

Since both ORC1a and ORC1b are expressed in the endocycle domain of the root we asked whether the dynamics in relation to S-phase was conserved in these cells, in spite that the labeling pattern was different in proliferating and endocycling cells. We found that in nuclei undergoing endocycle S-phase that were labeled with a 15 min of EdU pulse, both ORC1a (Fig. 28A) and ORC1b (Fig. 28B) were also excluded from most of the EdU positive endocycling cells. This shows that both proteins appear to be degraded also soon after entering the endocycle S-phase and synthesized again during the G-phase.

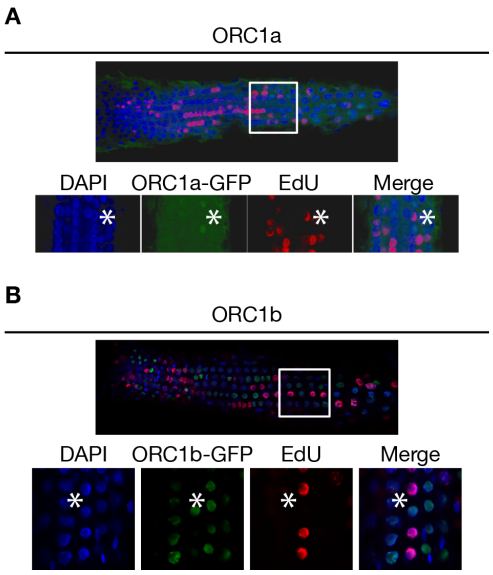


Figure 28 – S-phase endocycle nuclei are depleted of ORC1 proteins. S-phase cells were labeled with EdU (red) of plants expressing **(A)** ORC1a-GFP or **(B)** ORC1b-GFP (green). Nuclei were stained with DAPI (blue). Asterisks mark cells positive for EdU (S-phase) that lacks ORC1 proteins.

Differential localization of preRC proteins in Arabidopsis root

To study the expression of the ORC1 proteins in context with other pre-RC proteins, we evaluate the expression of the available translational constructs that express several pre-RC proteins fused to a GFP tag (Fig 29). Remarkably, they behave quite differently. Only ORC1a, ORC1b and ORC2, out of the proteins analyzed, showed a clear accumulation at chromocenters. Expression levels were different in proliferating and endocycling cells for all the pre-RC proteins studied except for ORC6, which GFP levels were more constant. MCM7 exhibits expression at the nucleus and the cytoplasm, probably due to the regulation mechanism of the protein. GFP-labeled mitotic figures were found in ORC1b and ORC2 expressing plants, but not in ORC6, CDT1a or MCM.

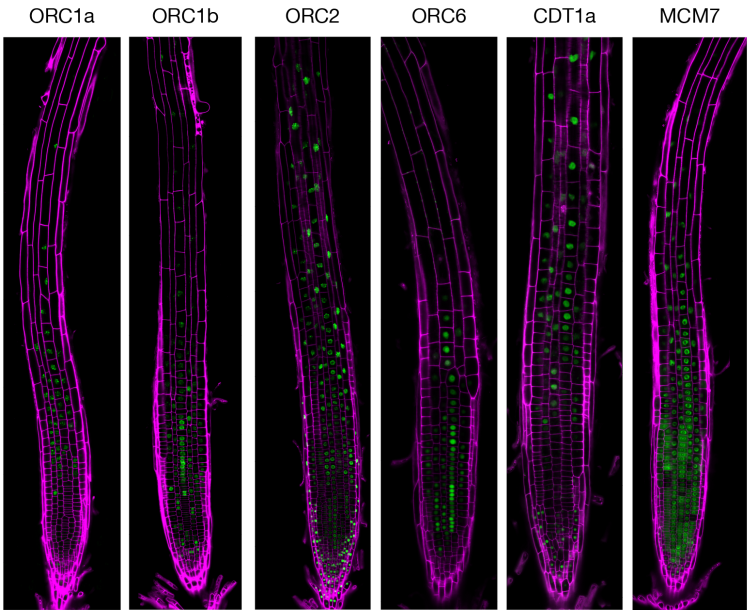


Figure 29 – Expression of pre-RC proteins at the root meristem. Plants expressing ORC1a, ORC1b, ORC2 (Ngo et al., 2012), ORC6 (Diaz-Trivino, 2005), CDT1a (Lopez et al., unpublished) and MCM7 (Herridge et al., 2014) fused to GFP were stained with propidium iodide and image under the confocal microscope.

4.2. Role of ORC1a and ORC1b in genome stability and heterochromatin maintenance

Identification of mutants in *ORC1* genes

The differences in expression domains, subcellular localization, cell cycle dynamics and protein stability of ORC1a and ORC1b strongly pointed to distinct roles of these proteins. To test this hypothesis we identified T-DNA insertion mutant alleles, *orc1a-1* and *orc1b-1*, (Fig. 30A). mRNA from 7dps whole seedling extracts was evaluated using several specific primers, as depicted in Figure 30A and Supplementary Table 3. We could detect the amplification using primers designed before and after the insertion of wild type cDNA but not for each mutant, and the amplification was positive for the primers before and inside the insertion in the mutants but not in the wild type plants (Fig. 30B). This result indicates that the wild type full length mRNAs are not being produced in the mutants. We have developed a specific antibody against ORC1b that recognizes a peptide from the N-term part of the protein (GMNLIRKRERAPR from residue 72 to 85). To evaluate if a truncated version of the protein ORC1b was produced in *orc1b-1* mutant plants we did WB assays using nuclear protein and this custom antibody (Fig. 30C). While in wild type plants the full-length protein is clearly detected, in the mutant plants the band is not present, suggesting that the full-length functional protein is not produced.

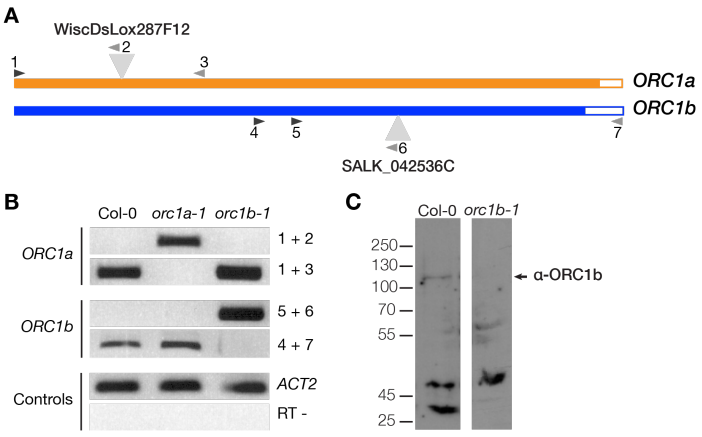


Figure 30 – *ORC1* expression levels in mutants. (A) Schematic representation of *ORC1a* (orange) and *ORC1b* (blue) genes, the insertion points of the T-DNAs (grey triangles) and the primers used to evaluate the mRNA levels are indicated in grey. **(B)** mRNA expression around the T-DNA insertion in *ORC1* mutant and Col-0 wild type plants. The expression of *ACT2* (*ACTIN2* AT3G18780) was monitored as a positive control and a minus reverse transcriptase (RT-) as a negative control. Primer sequences are listed in Supplementary Table 3. **(C)** Western blot assay showing wild type (Col-0) and *orc1b-1* nuclear protein extracts. The ORC1b protein was detected with a specific antibody.

We also crossed *orc1a-1* and *orc1b-1* mutants to study the effect of the depletion of both ORC1 proteins. Unfortunately, out of 88 plants from the F2 generation checked, none of them was homozygous for both mutations. The two *ORC1* genes are in the long arm of chromosome 4, 965 Kb apart. It has been reported that the recombination frequency in this chromosome is very low (Drouaud et al., 2006; Drouaud et al., 2007) compared to other Arabidopsis chromosomes. All together the results suggested that *ORC1a* and *ORC1b* genes are in linkage disequilibrium.

The primary root of *orc1* mutants presents a normal growth

Mutations in the *ORC1* genes did not produce drastic macroscopic phenotypes, as deduced from the rather normal appearance of plants at the vegetative and reproductive stages. Since the expression of these proteins is confined to the meristems, we investigated in more detail features directly related to cell proliferation within the meristems. We assessed the growth of the RAM organization by measuring the length and number of cortical cells from the QC to the elongation zone. These mutations did not have significant effects on RAM size, neither at 4 dps when it is immature, nor at 7 dps (Fig. 31A) when it is fully develop. Thus, meristem cortical cell number (Fig. 31B) and meristem size (Fig. 31C) was not significantly different between wild type and mutant roots.

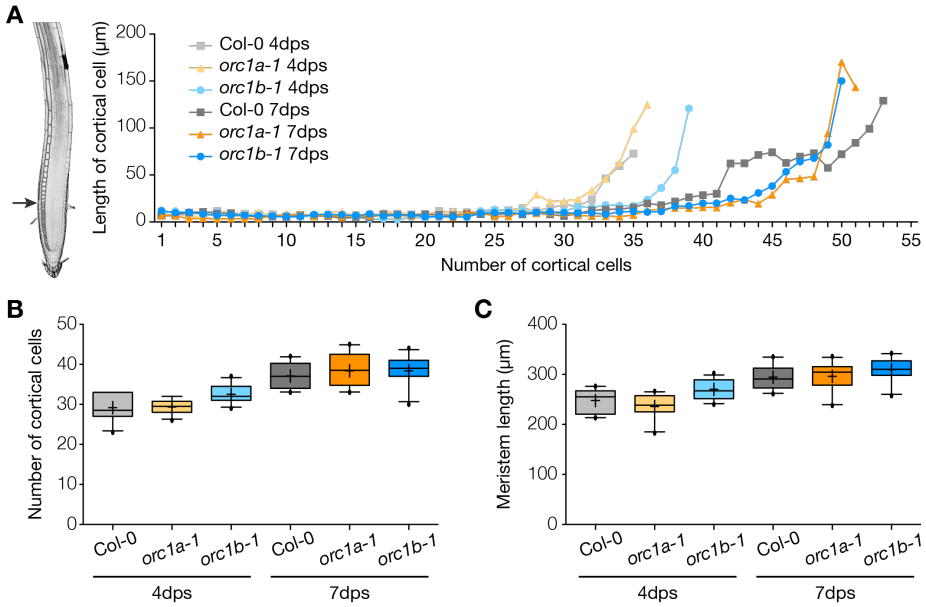


Figure 31 –Size of RAM in *ORC1* mutants. (A) Cortical cell length from the QC to the elongation zone of Col-0 wild type (grey), *orc1a-1* (orange), and *orc1b-1* (blue) at 4 dps or 7 dps. At the left a representation of the RAM is showed. **(B)** Meristem cortical cell number in the three genotypes at 4 dps and 7 dps. The distribution from the 10 to the 90 percentile is plotted. **(D)** Meristem length for the three genotypes at 4 dps and 7 dps. The distribution from the 10 to the 90 percentile is plotted. For **(B)** and **(C)** non-significant differences were found between mutant and the wild type plants in a one-way ANOVA analysis using a Kruskal-Wallis test and Dunn's multiple comparisons. $n_{4\ dps}$ = 12 roots, $n_{7\ dps}$ = 10 roots.

Furthermore, no major changes were detected in the 2C, 4C, 8C and 16C cell populations in the root. We analyzed the nuclei from the tip of the 7 dps roots by flow cytometry and found that the distribution of the cell populations with different DNA contents was not drastically modified in the wild type and *orc1* mutants (Fig. 32).

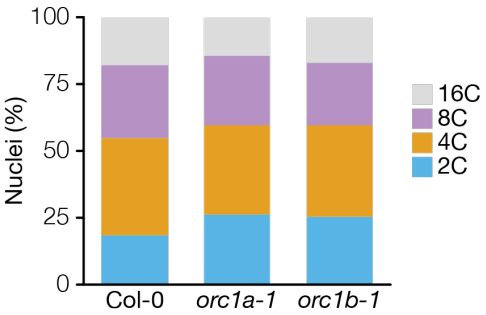


Figure 32 – DNA content of cells is maintained in *orc1* mutant roots. The tip of 7dps roots from wild type (Col-0), *orc1a-1* and *orc1b-1* roots was chopped in Galbraith buffer to isolate nuclei and stained with DAPI prior to flow cytometry evaluation. The four populations of cells (2C –blue–, 4C –orange– , 8C –green– and 16C –pink–) are unaltered in *orc1* mutants.

Therefore, and somehow surprisingly, our studies of the RAM suggested that the depletion of ORC1a or ORC1b in the roots do not produce a severe growth phenotype under normal conditions.

orc1 mutants present a delay in S-phase progression

The canonical role of ORC1 proteins is to participate with the rest of ORC proteins in pre-RC formation at the sites were DNA replication could be initiated. To monitor DNA replication *in vivo* in the growing root we labeled the nuclei in S-phase with a pulse of 15 min pulse of EdU of wild type and *orc1* mutant roots. We found an increase in the number of cells labeled with EdU from 29.3% in the wild type to 32.4 and 34.3% in *orc1a-1* and *orc1b-1*, respectively (Fig.33). Although the differences were not statistically significant, the result pointed to a defect in DNA replication. The analysis of the EdU patterns corresponding to early, mid and late S-phase, according to the position of the nuclei along the root meristem, indicated that early S-phase nuclei from *orc1b-1* were closer to the QC and the proliferation zone compare to the wild type Col-0 (Fig. 33). This indicates a defect in the early replicating nuclei in the *orc1b-1* mutant plant.

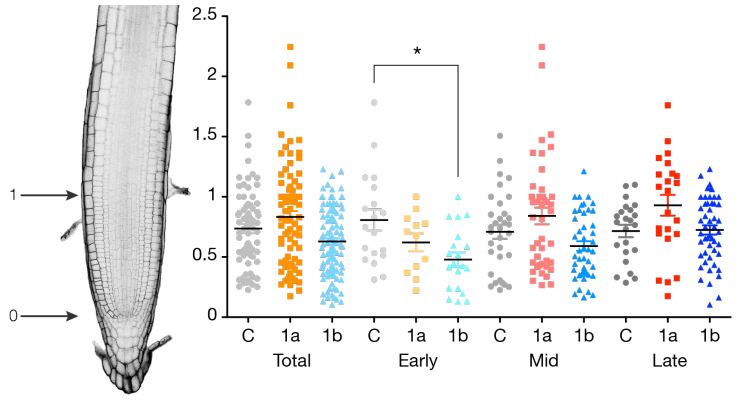


Figure 33 – Early S-phase cells in orc1b-1 are closer to the QC. Roots from wild type Col-0 (C), *orc1a-1* (1a) and *orc1b-1* (1b) mutants were labeled with a 15 min EdU pulse. Nuclei undergoing S-phase were positioned along the root (Total) and separated into early, mid and late S-phase.

We then followed S-phase progression using a double pulse-labeling strategy. We first labeled cells with BrdU for 15 min, then incubated the roots with thymidine letting S-phase to progress and finally labeled with EdU for another 15 min. This allowed us to estimate S-phase length as the chase time required to obtain no double-labeled cells. When there is no time

between pulses both signals from BrdU (yellow) and EdU (red) are detected colocalizing in the same nuclei and largely at the same nuclear region (Fig. 34A). Quantification of the percentage of cells labeled with EdU and BrdU revealed a progressive decrease in the amount of double-labeled cells with increasing chase times, as expected (Fig. 4B) as well as allowed an estimation of the S-phase length of 2-2.5h in these cells.

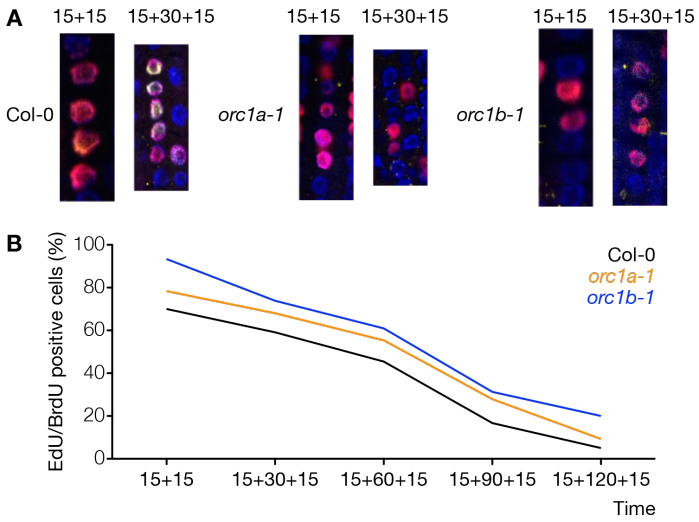


Figure 34 – S-phase progression is delay in ORC1 mutants. (A) Examples of the double-labeled nuclei in wild type and *orc1* mutants. **(B)** Quantification of the progressive decrease of double-labeled nuclei.

In the *orc1* mutants, reaching the state where the two labels occurred in different nuclei took more time than in the wild type (Fig. 34A, B). In fact, in the first time point the difference between the percentage of double-labeled cells in the wild type and in *orc1b-1* mutant suggests that the progression in early S-phase nuclei was slower in the mutant. Remarkably, the progression in *orc1b-1* mutants was slower than in the *orc1a-1* mutant, in which it was almost undistinguishable from the wild type.

Depletion of ORC1b blocks growth upon aphidicolin stress conditions

To test whether the impairment of *ORC1* mutants to progress through S-phase was enhanced under DNA replication stress conditions, we treated the plants with different drugs that produce DNA damage or affect fork progression. We evaluated the growth of the roots in the presence of zeocin (that produces double strand breaks), hydroxyurea (that slows down the forks due to a depletion in the dNTP pull), and aphidicolin (that blocks DNA polymerase α and fork progression). We observed that these treatments had very different consequences, as summarized in Fig. 35A. Thus, while the sensitivity of *orc1a-1* to these treatments was similar to that of the wild type, aphidicolin produced a severe growth arrest of the *orc1b-1* mutant (Fig. 35A, lower righter panel, 35B).

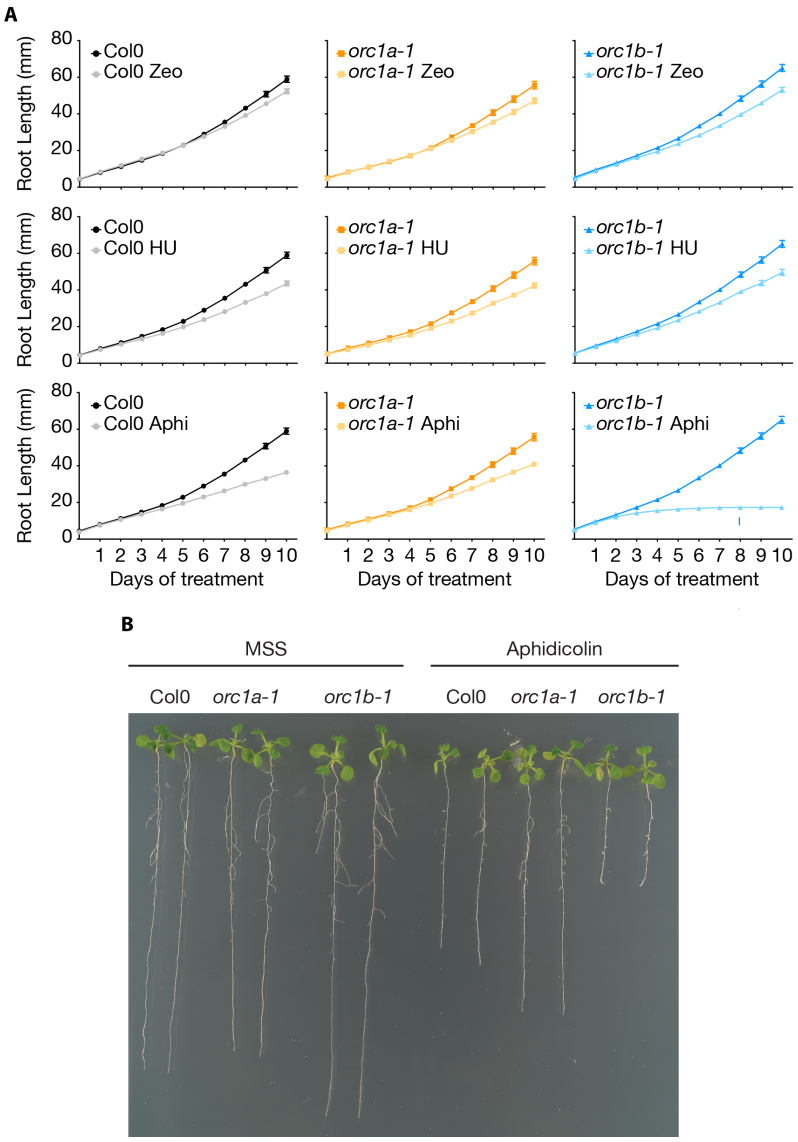


Figure 35 – *orc1b-1* root growth is blocked under aphidicolin conditions. (A) 3 dps plants were transferred to plates containing either 6.7 µM Zeocin (Zeo), 1 mM hydroxyurea (HU) or 0.12 µg/mL aphidicolin (Aphi). Root length was evaluated every 24 h for a total of 10 days. The mean and the standard error of the mean are plotted for 2 biological experiments, each one with 36 plants per genotype. **(B)** After 10 days of treatment with aphidicolin, *orc1b-1* root growth is severely impaired compare to the wild type and *orc1a-1* mutant plants, while in control medium (MSS) the differences are not significant.

The different effect between HU and aphidicolin treatment was unexpected. To assess if the differences were due to the drug concentration we treated the plants with 5 and 10 mM HU (Fig. 36). Using higher doses of HU led to the death of the plants (wild type and both *orc1* mutants), reinforcing the idea that *orc1* mutants tolerate HU stress like wild type plants.

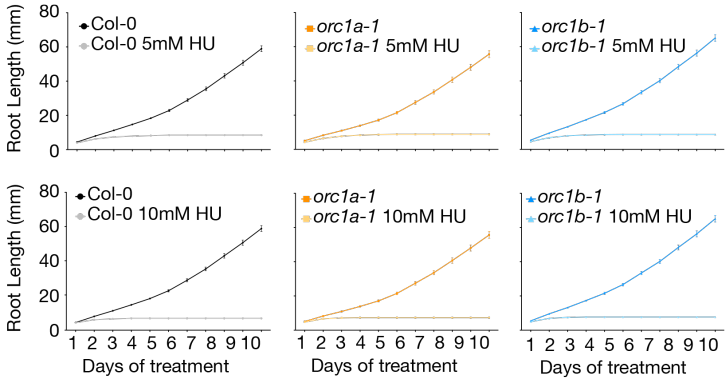


Figure 36 – orc1 mutants tolerate HU stress. 3 dps plants were transferred to plates containing either 5 or 10 mM hydroxyurea (HU). Root length was evaluated every 24 h for a total of 10 days. The mean and the standard error of the mean are plotted for 2 biological experiments, each one with 36 plants per genotype.

The hypersensitivity of *orc1b-1* mutant to aphidicolin was not due to defects in triggering the G2 DNA damage checkpoint as revealed by (i) the tolerance to zeocin and (ii) the upregulation of the mRNA of *RAD51*, *BRCA1* and *PARP1* genes, which occurred to levels comparable to those in the wild type or *orc1a*-1 mutant (Fig. 37).

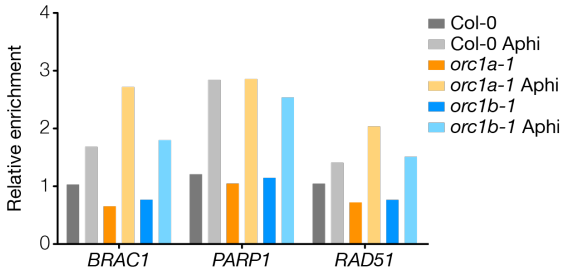


Figure 37 – G2 Checkpoint related genes are active in orc1b-1 upon aphidicolin stress. mRNA was isolated from whole seedlings after 10 days of treatment with 0.12 µg/mL aphidicolin. Expression levels were normalized against the reference gene *GAPDC2 (GLICERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2)* and wild type in control conditions.

These results suggest that the hypersensitivity of *orc1b-1* to aphidicolin was due to problems in pre-RC licensing and ORI firing.

The absence of ORC1a leads to defects in heterochromatin maintenance

In yeast and animals it has been reported an additional function of Orc1 protein in heterochromatin formation and maintenance (Chakraborty et al., 2011; Prasanth et al., 2010). We have found that ORC1b, which is present in proliferating and endocycling cells, seems to be involved in replicative functions both in euchromatin and heterochromatin. In contrast, our results indicate that the primary role of ORC1a does not seem to be related to DNA replication. Instead, based on the specific localization of ORC1a at the chromocenters in endocycling cells at the transition to elongation zone of the root, it could play a role in heterochromatin formation and maintenance.

To test this possibility we evaluated, using immunofluorescence microscopy, the level of two typical plant heterochromatin marks, H3K9me2 (deposited by KYP, SUVH5 and SUVH6) and H3K27me1 (by ATXR5 and ATXR6 (Bernatavichute et al., 2008; Du et al., 2015; Jacob et al., 2009)). We first evaluated the proportion of both marks in nuclei at the transition and elongation zone of the root and found that in *orc1a-1* mutant there was a subpopulation of nuclei with high amounts of H3K9me2 and low amounts of H3K27me1 (Fig 38A). Remarkably, this was not the situation for *orc1b-1* mutants (Fig. 38B) and the situation resembled that of the *atxr5/6* double mutant (Fig. 38C), although less apparent than for *orc1a-1*. The *orc1a-1* endocycling nuclei are distributed in two populations, one with normal H3K27me1, relative to H3K9me2, and another with lower amounts of H3K27me1 compared to H3K9me2. This result clearly suggests a previously unidentified role of ORC1a in the H3K27me1 deposition.

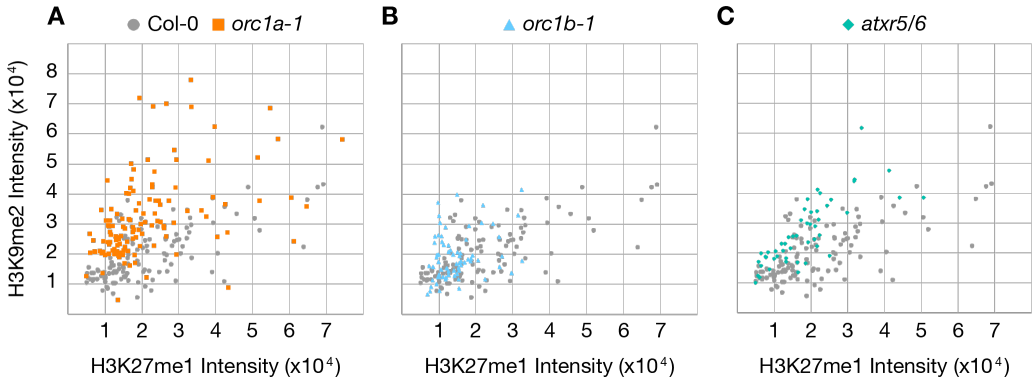


Figure 38 - H3K27me1 levels distinguished two populations of endocycle cells in *orc1a-1* mutant roots. H3K9me2 and H3K27me1 marks were immunodetected in 7 dps roots of Col-0 wild type (grey), *orc1a-1* (A, orange), *orc1b-1* (B, blue) and *atxr5/6* (C, green) mutants. The fluorescence level for each mark was measured in the nuclei of the transition to elongation zone of the root and plotted.

To gain a better insight into the defects in heterochromatic marks of these two populations, we evaluated the nuclei one by one. We found that in wild type roots the signals for H3K9me2 and H3K27me1 colocalized very nicely with chromocenters, as previously reported ((Naumann et al., 2005); Fig. 39A).

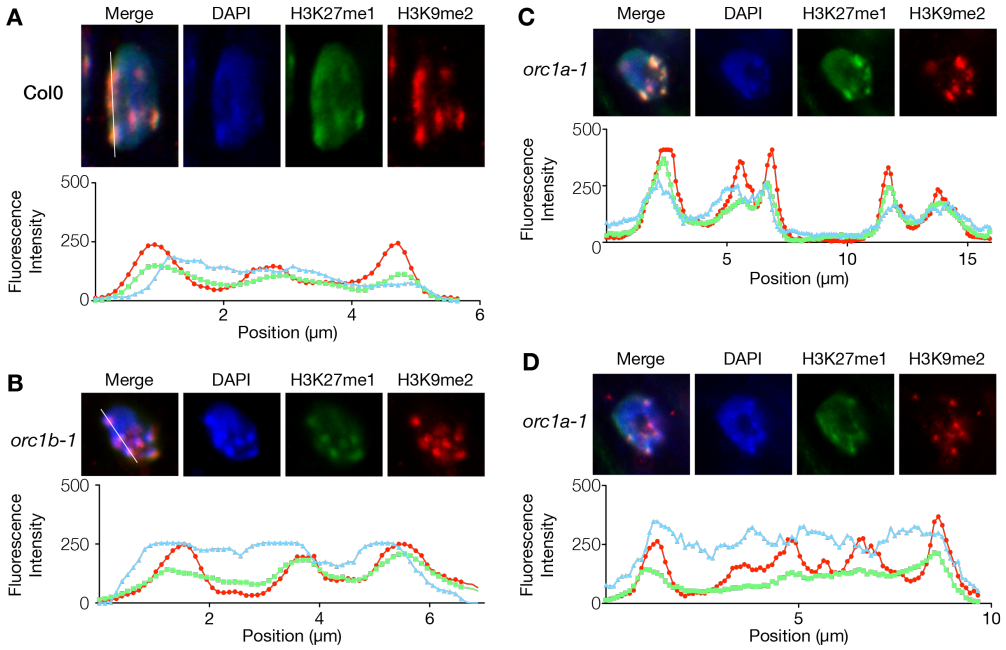


Figure 39 – H3K27me1 is decreased in the endocycle cells of *orc1a-1* mutant roots. Roots from 7 dps plants of **(A)** Col-0 wild type, **(B)** *orc1b-1* and **(C, D)** *orc1a-1* were immunostained to detect H3K9me2 (red) and H3K27me1 (green) and all nuclei were stained with DAPI (blue). The fluorescent levels for each mark was measured in epidermal nuclei of the endocycling cells at the transition zone. In *orc1a-1* two different populations were identified, one with normal levels of H3K27me1 **(C)** and other with lower levels **(D)**.

That was also the case for the nuclei of the *orc1b-1* mutant (Fig. 39B). Interestingly, we discovered that the lack of ORC1a had a significant effect in the level of H3K27me1 labeling at chromocenters. One of the nuclei population in the endocycle zone of *orc1a-1* mutant roots exhibited a wild type phenotype (Fig. 39C) while another present absence of H3K27me1 signal at chromocenters or the signal was severely diminished (Fig. 39D). This result suggests that ORC1a participates in chromocenter organization by determining the final levels of H3K27me1.

The *atxr5/6* double mutants present a strong reactivation of the TEs (Jacob et al., 2009) caused by the loss of H3K27me1. We evaluated the expression levels of a collection of TEs by qPCR and found that in both *orc1* mutants this collection of TEs remains silent (Fig. 40). It has to be kept in mind that the PCR analysis was carried out in the whole seedlings and the differences in H3K27me1 labeling were detected in a subset of endocycling cells of the root. Also, only a set of TEs has been evaluated. However, this result indicates that the decrease in H3K27me1 caused by the absence of ORC1a is not as severe as the observed in the *atxr5/6* mutant.

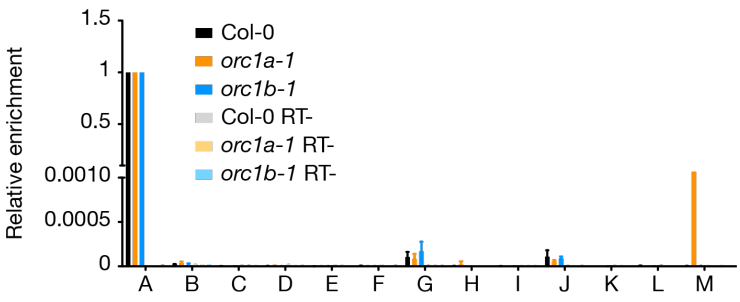


Figure 40 – TEs are silent in *orc1* mutant plants. mRNA was isolated from whole 7 dps plants of wild type Col-0, *orc1a-1* and *orc1b-1* mutants. Expression levels were normalized against the reference gene *GAPDC2 (GLICERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2)*. The codes for the primer pairs used to identify each TE, according to the list in Supplementary Table 4 are: A, GAPDC2 AT1G13440; B - AT2TE13970; C - AT2TE16335; D - AT4TE16735; E - AT4TE17050; F - AT4TE16725-2; G - AT4TE16725-3; H - AT2TE15565-2; I - AT2TE15565-3; J - AT1TE62820-3; K - AT1TE62820-5; L - AT4TE03295-3; M - AT4TE03295-4. M – AT4TE37370. RT- = Minus reverse transcriptase control.

4.3. DNA replication origins colocalize with retrotransposons at pericentromeric regions

ORI specification of euchromatic chromosome domains is relatively known in various model systems and has been related to marks of open and active chromatin. In contrast, the genomic features that contribute to specify ORIs in heterochromatin have not been investigated in detail and, consequently, are very poorly understood. Therefore, we sought to elucidate the genomic features defining ORI localization in Arabidopsis heterochromatin, which is largely concentrated in the pericentromeric regions.

Mapping of ORIs at the pericentromeric heterochromatin

The identification of ORIs responsible for replication of pericentromeric heterochromatin requires very reliable genome annotation and peak calling algorithms. In the case of *Arabidopsis thaliana*, an updated genome annotation (TAIR10), including highly repetitive pericentromeric regions, is available. We have used the sequencing data of purified BrdU-pulsed labeled DNA isolated from Arabidopsis cultured cells synchronized at early S-phase (GSE21828, (Costas et al., 2011)) to generate a high-resolution map of ORIs, paying particular attention to those located in heterochromatin domains. Since heterochromatic regions are packed with repetitive elements such as TEs or satellites, we first discarded the multihit reads for the alignment, in other words, the reads that align in more than one location in the genome. Discarding multihit reads rendered a more confident set of ORIs on those regions, although it leads to an underestimation of the number of reads mapping to repetitive elements. Since there is no information about the characteristics of ORIs at heterochromatic sites, we decided to use six different peak calling algorithms, namely, MACS (Zhang et al., 2008); versions 1.4 and 2.0), BayesPeak (Spyrou et al., 2009), T-PIC (Hower et al., 2011), HOMER (Heinz et al., 2010) and SICER (Zang et al., 2009). Although the distribution of the peak size was different for each peak calling, all six of them rendered small peaks (Fig. 41A). The mean size ranged from 537 to 1232, corresponding to MACS2.0 and T-PIC, respectively. Additionally, the peak size distribution was not different between the pericentromeric (heterochromatin) and non-pericentromeric regions (Fig. 41A). While MACS2.0, BayesPeak and HOMER failed to detect clear peaks, T-PIC and SICER produced false positive peaks at noise regions (Fig. 41B). Finally, after visual inspection of the peaks in a genome browser, we selected MACS1.4 as the algorithm best matching our raw sequencing data.

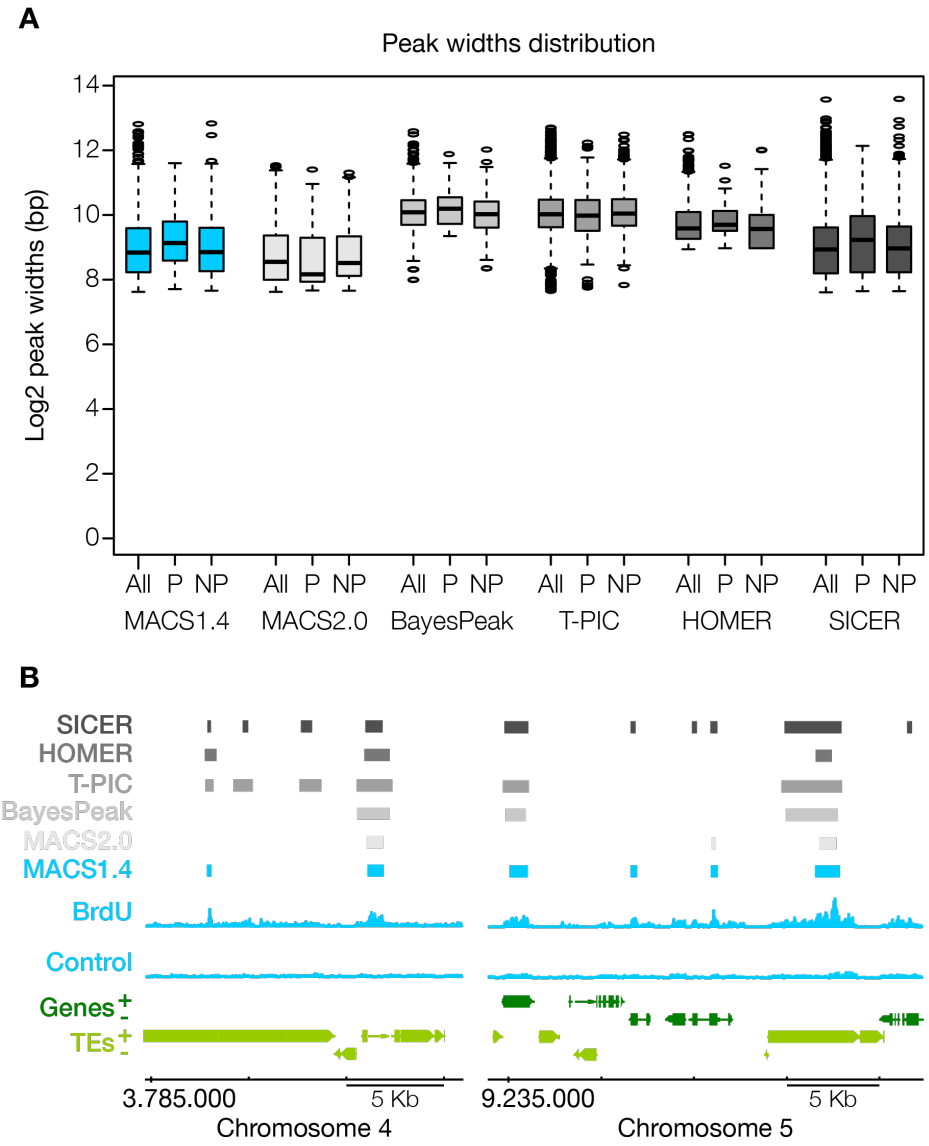


Figure 41 – Comparison between peak calling algorithms. (A) Boxplot distribution of the peak size obtained from six peak calling algorithms at All: all the genome, P: pericentromeric and NP: non-pericentromeric regions. **(B)** Representative genome-browser views of regions containing ORIs of chromosomes 4 and 5, as indicated. BrdU and control reads, as well as the selected MACS1.4 peak calling is shown in blue. Genes (dark green) transcribed from each strand and TEs (light green) are shown along the chromosome together with the coordinate scale.

The analysis using MACS1.4 showed that ORIs have a strong preference to colocalize with genes. Out of a total of 3230 ORIs in the entire genome, 2888 (89.4%) colocalized with genes and 161 (4.9%) with TEs (Fig. 42A), a result in accordance with previous overall analyses (Costas et al., 2011). However, our analysis also showed that the proportions change drastically when we consider separately non-pericentromeric (chromosome arms) and pericentromeric regions. Indeed, whereas almost all ORIs (94.9%) colocalize with genes in gene-rich domains of chromosome arms, less than half of ORIs (46.7%) colocalize with genes,

and 33.7% with genes in the pericentromeric gene-poor regions (Fig. 42B,C). Furthermore, the distribution of ORIs located outside genes positively correlates with the distribution of TEs, and not with the distribution of non-annotated regions (Fig. 42D). Analysis of ORI-TE density along the Arabidopsis chromosomes visualizes the preference of non-genic ORIs to colocalize with TEs in pericentromeric regions (Fig. 42E). These results suggest that TE sequences may be selected as ORIs in regions with a low gene density such as pericentromeric regions.

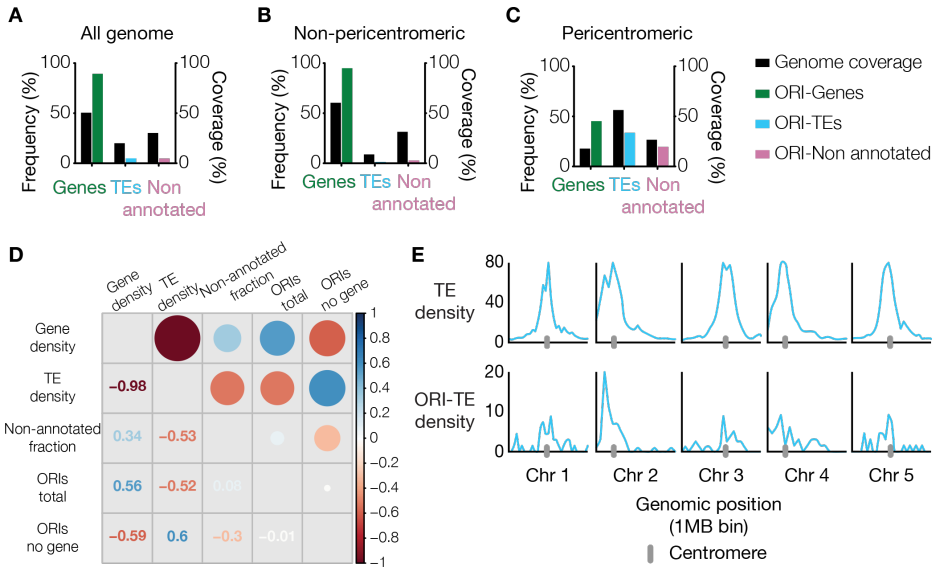


Figure 42 - Genomic location of Arabidopsis DNA replication origins. Fraction of ORIs found in genes, TEs and non-annotated regions in **(A)** all the Arabidopsis genome, **(B)** the non-pericentromeric regions and **(C)** the pericentromeric regions, defined as having a gene frequency ≤40%, shown with the respective genome coverage. **(D)** Overall correlation between gene, TE and non-annotated fraction coverage and total ORIs and ORIs not located in genes. Correlations are represented with circles (gradation of red, anticorrelation; gradation of blue, positive correlation). The size of the circles corresponds to the correlation coeficient, also indicated in the other half of the plot. **(E)** TE density (% of nucleotides in TEs per 1 Mb bin) (upper panels) and chromosomal distributions of ORI-TEs across the five Arabidopsis chromosomes (lower panels).

To evaluate if the distribution of ORIs in TEs was affected by choosing the uniquely mapped reads, we repeated the analysis taking into account only the multihit sequence reads (Table 1). We found 246 multihit-ORIs, and only 77 (31.3%) colocalize with genes. However, the number of ORIs colocalizing with TEs and non-annotated regions increased, 48% and 19.1%, respectively. This was probably due to the higher abundance of repetitive sequences at this type of elements.

Feature	All genome	
	No. ORIs	% ORIs
Genes	77	31.3
TEs	118	48
Non-annotated	47	19.1
Overlapping	4	1.6
Total	246	100

Table 1 – Distribution of ORIs from multihit reads. Fraction of ORIs resulting from the analysis of multihit reads found in genes, TEs and non-annotated regions in all the genome.

ORI-TEs preferentially colocalize with retrotransposons

TEs constitute a very heterogeneous type of repetitive elements. They can be divided mainly in two classes: retrotransposons and DNA transposons. Thus, we first asked whether ORIs in TEs were homogenously distributed among the various TE families and found a striking preference for ORIs to associate with certain TE families (Fig. 43). Very surprisingly the two most common families that occupy 29.4% of the total TE coverage each, the retrotransposons Gypsy and the DNA transposons Helitron, behave completely opposite regarding the presence of ORIs. While Gypsy elements contain most of the ORI-TEs (46.98%), Helitron elements completely lack ORIs (Fig. 43A). This significant difference is maintained in both the non-pericentromeric (Fig. 3B) and pericentromeric (Fig. 3C) regions. We have found that other DNA transposon families such as DNA/MuDR elements contain very little ORIs, only 3.36% of the total ORI-TEs. In contrast, retrotransposon families such as the already mentioned Gypsy or LINE and Copia elements contain a high amount of ORIs (22.82% and 14.09% respectively).

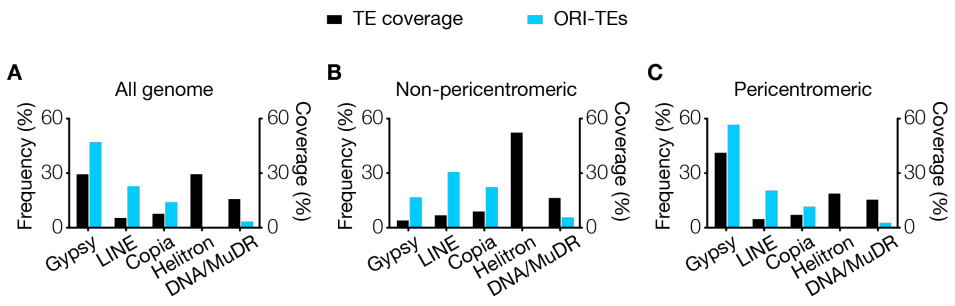


Figure 43 – Frequency distribution of ORI-TEs in TE families. (A) All the Arabidopsis genome. **(B)** Non-pericentromeric regions. **(C)** Pericentromeric regions, (blue bar) shown with the respective TE family nucleotide coverage of total TE nucleotides (black bar).

Noteworthy, the distribution of ORI-TEs obtained from various peak calling algorithms lead to the same general conclusion (Fig. 44A). Only T-PIC rendered some ORIs (1.32%) within the Helitron elements, although most of the ORI-TEs colocalize with retrotransposons.

Furthermore, analysis of the multihit sequences revealed similar results (Fig. 44B), indicating that the lack of ORIs in Helitrons is not due to a bias derived from sequence alignment problems. Together, these observations demonstrate that when ORIs associate with TEs they have a significant preference to colocalize with retrotransposons and specifically Gypsy elements, whereas they tend to be excluded from DNA transposons, in particular from Helitron elements.

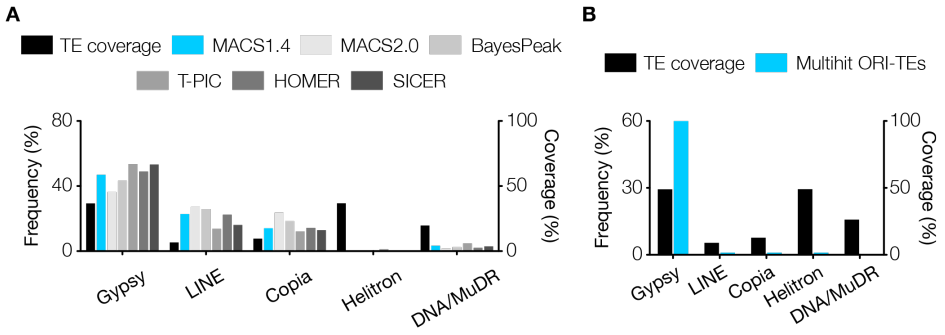


Figure 44 – Frequency distribution of ORI-TEs in TE families. Distribution of ORI-TEs along the TE families **(A)** derived from six peak callings algorithms, namely, MACS1.4 (blue bars), MACS2.0, BayesPeak, T-PIC, HOMER, and SICER (grey bars)or **(B)** calculated using the multihit sequence reads (blue bars) and the corresponding TE coverage (black bars).

Short nascent DNA strands (SNS) enrichment confirms the activity of ORIs mapped by BrdU-seq

To validate our ORI mapping strategy using an independent method we determined the activity of a number of ORIs by quantitative PCR enrichment of a purified sample of short nascent strands (SNS). For a detailed validation of ORI activity we designed sets of primer pairs across a chromosomal region containing one ORI overlapping with a TE in the arm of chromosome 1 (AT1TE62820) and another ORI ~70 kb apart, colocalizing with a downstream gene (AT1G51350) within a typical euchromatic region. Cultured Arabidopsis cells were synchronized in G0 by sucrose deprivation and then samples were extracted 2 (G1/S), 3.5 (early S) and 7 h (late S) after release from the sucrose block. qPCR analysis was carried out in three biological replicates and in two consecutive fractions of the sucrose gradient to ensure reproducibility of the data. As expected, none of the ORIs selected were active at the earliest time point analyzed, 2h after release of the sucrose block (Fig. 45A). At later time points, a clear enrichment was detected in both cases, revealing the activity of these two ORIs in the cell population. Also, it is worth noting that the ORI located within a gene (Fig. 45A, right panels) was ~5-10-fold more active than the ORI colocalizing with a TE (Fig. 45A, left panels). These experiments confirm that both predicted ORIs, located in a TE and in a gene, indeed function

as ORIs. This analysis also showed that an ORI located at a TE in a chromosome arm is active in cultured cells, even when another stronger ORI is in the neighborhood, less than ~70 kb apart.

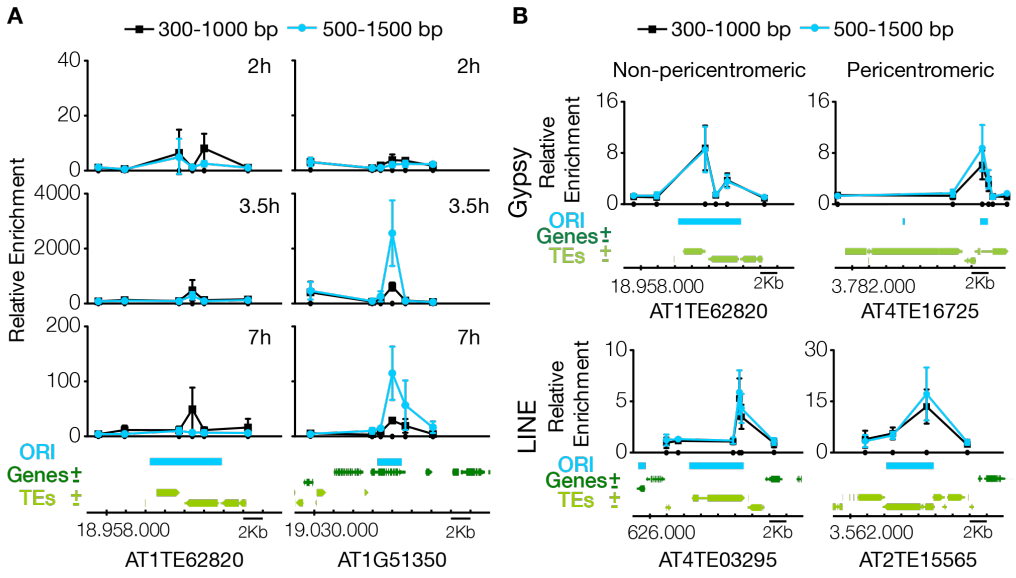


Figure 45 – DNA replication origin activity determined by short nascent strand (SNS) abundance by qPCR. (A) Measurement of ORI activity in synchronized Arabidopsis MM2d cells at various times after releasing the block, as indicated (2h, G1/S; 3.5h, S; 7h, late S). In each case, the confidence of ORI activity was assessed by analyzing in three biological replicates two consecutive fractions, as indicated at the top. The fractions belong to the same gradient used for purification of SNS and contain DNA molecules ranging 300-1500 bp in size. Two ORI-containing regions (left panels, ORI colocalizing with a TE; right panels, ORI colocalizing with a neighbor gene) were analyzed. The location of primer pairs scanning the region is indicated by small dots on the X-axis. Enrichment values were made relative to the flanking region and normalized against gDNA. The genomic region under study depicting the location of ORI, genes and TEs is at the bottom. Chromosomal coordinates are indicated. **(B)** Measurement of ORI activity in asynchronous Arabidopsis MM2d cell cultures. The ORI-TEs were chosen according to their family (Gypsy and LINE) and location (non-peri- and pericentromeric), as indicated. The location of primer pairs is indicated by small dots on the X-axis. Two consecutive fractions were analyzed, as described for panel A. Enrichment values were made relative to a negative region that does not content any ORI or TE (AT2G28970). The genomic region under study depicting the location of ORI, genes and TEs is at the bottom. Chromosomal coordinates are indicated

We also wanted to evaluate the activity of different ORI-TEs according to the TE family they colocalize with. Thus, we chose to validate and analyze in asynchronous cells, four genomic regions containing ORI-TEs: two belonging to the Gypsy family and two belonging to the LINE family (where ORIs are highly and moderately over-represented, respectively), and in each case one ORI located in pericentromeric heterochromatin and another in non-pericentromeric heterochromatic patches within the euchromatic arms. We found that all ORI-TEs analyzed here were active as revealed by the qPCR enrichment of purified SNS (Fig. 45B). These experiments confirm that the results obtained by direct sequence mapping of BrdU-

labeled material represents a *bona fide* collection of active ORIs in heterochromatin and that TEs are a major source of ORIs in pericentromeric regions.

ORI-TEs activity occurs independently of gene expression

ORI activity has been related with active gene transcription (Mechali et al., 2013; Sequeira-Mendes et al., 2009). Although the expression of TEs is usually strongly repressed, some TEs can be activated under stress situations (Deragon et al., 2008; Lisch, 2013). Notably, it was reported that in an Arabidopsis cell culture line, typical heterochromatin marks change and some TEs are activated (Tanurdzic et al., 2008). In order to verify if the presence of ORIs within retrotransposon TEs was due to a reactivation of those elements we isolated total RNA and analyze it using either oligodT or random hexameres to perform the retrotranscription. In all the cases RNA levels were below the detection threshold (Fig. 46), neither the corresponding to retrotransposons nor in Helitron elements nearby the analyzed pericentromeric TEs. Consequently, we concluded that ORI-TE activity in our Arabidopsis cell culture line is independent of the transcriptional status of the TEs they are associated with. Based on these observations, we sought to identify whether a unique signature can be associated with the high preference of retrotransposon families for ORI specification.

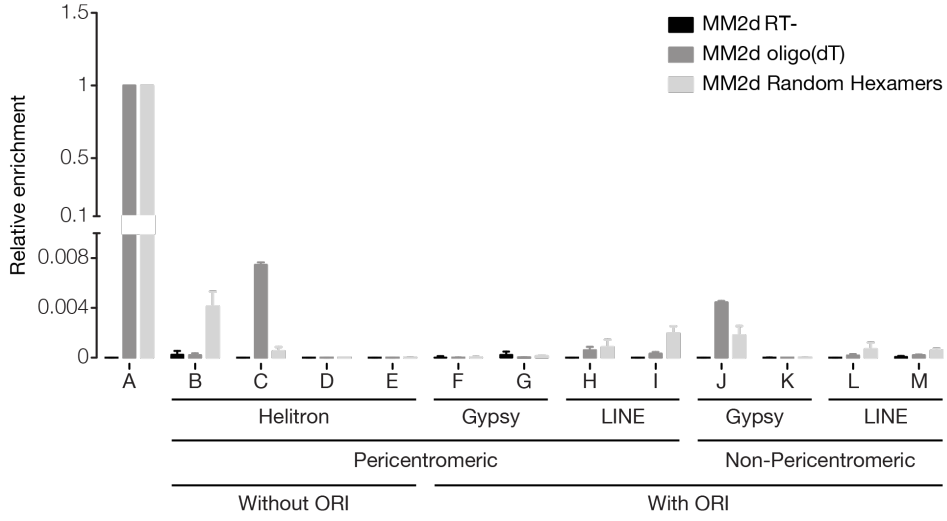


Figure 46 – RNA levels in Arabidopsis MM2d cultured cells across TEs representative of various families, with or without ORIs. Enrichment values were calculated as described in Methods and normalized against the reference gene *GAPDC2 (GLICERALDEHYDE-3-PHOSPHATE DEHYDROGENASE C-2).* The codes for the primer pairs used to identify each TE, according to the list in Supplementary Table 4 are: A, GAPDC2 AT1G13440; B - AT2TE13970; C - AT2TE16335; D - AT4TE16735; E - AT4TE17050; F - AT4TE16725-2; G - AT4TE16725-3; H - AT2TE15565-2; I - AT2TE15565-3; J - AT1TE62820-3; K - AT1TE62820-5; L - AT4TE03295-3; M - AT4TE03295-4. RT- = Minus reverse transcriptase control.

The activity of ORI-TEs is maintained with high levels of mC and is independent of G quadruplexes.

The majority of ORIs colocalize with genes which, when highly expressed, tend to be highly methylated at CG positions within the gene body, but not at CHG or CHH, the other sequence contexts where C methylation is found in plants (Zhang et al., 2006). Moreover, the ±100 nt region around the ORI in euchromatin tends to be depleted of CG methylation (Costas et al., 2011), which suggests that ORI specification and activity may depend on low levels of methylation. TEs are heavily methylated in C residues of the three sequence contexts, and their methylation is actively maintained by RNA-directed DNA methylation (RdDM) and siRNAs (Fultz et al., 2015; Matzke and Mosher, 2014). However, TEs may differ in their methylation state depending on the type, size or location (Ahmed et al., 2011; Zemach et al., 2013). Thus, we used the available methylation data of the Arabidopsis genome (Stroud et al., 2013) to ask whether differences in C methylation correlate with the preferential location of ORIs in certain TE family members. We found a tendency of Helitron elements, which do not colocalize with ORIs, to contain lower levels of C methylation for the three sequence contexts, whereas Gypsy elements, the most ORI-enriched TEs, showed higher methylation level (Fig. 47A). This is in line with previous reports that showed that Helitrons tend to be less heavily methylated than Gypsy elements in Arabidopsis (Ahmed et al., 2011). Moreover, the level of C methylation of Gypsy elements does not vary depending on whether they colocalize or not with ORIs (Fig. 47B). Therefore, our data suggest that a low methylation level is not a requirement for ORI specification in TEs. Similar observations have been made for the heterochromatic X chromosome in mammalian cells where the level of C methylation does not affect ORI specification and usage (Gomez and Brockdorff, 2004).

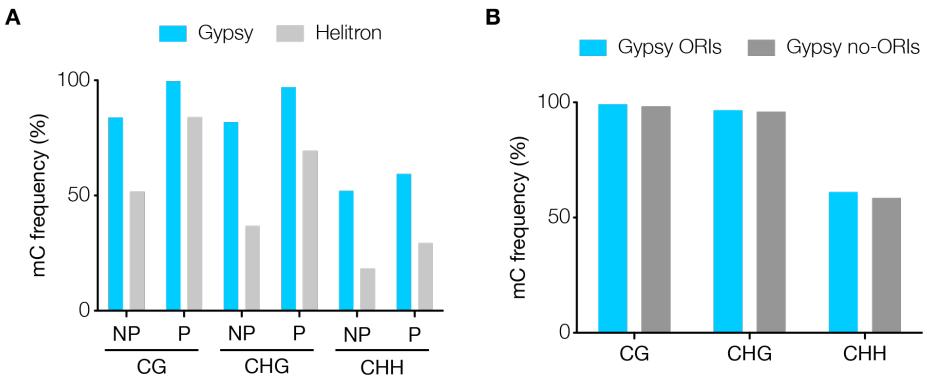


Figure 47 - Fraction of cytosine methylation. Fraction of methylated cytosines for **(A)** Gypsy (blue) and Helitron (light grey) elements, or **(B)** Gypsy colocalizing with ORIs (blue) or without them (dark grey), depending on the context (CG, CHG, CHH, where H=A, T or C) and region (P, pericentromeric; NP, non-pericentromeric).

G quadruplexes (G4) have been frequently found in association with TEs (Kejnovsky et al., 2015) and with ORIs in mammalian cells (Besnard et al., 2012; Cayrou et al., 2012; Comoglio et al., 2015; Valton et al., 2014). Thus, we also asked whether the presence of G4 was a determinant factor in the distribution of ORIs in Arabidopsis cells. We found first that G4 motifs are far more frequent in TEs than in genes whereas ORIs highly prefer a colocalization with genes. Second, most G4 motifs occur in a TE family known as ATREP18, which contains a canonical telomeric repeat (Cardenas et al., 2012) and that is also found in pericentromeric regions. This family is included within the annotation class "DNA/Other" that contains less than ~1% of all ORI-TEs (Fig. 48). Third, and perhaps more relevant, both Gypsy and Helitron elements contain a very similar fraction of G4 motifs whereas they show an opposite preference to contain ORI-TEs (Fig. 48). Hence, our observations do not support the idea that G4 structures may be directly influencing ORI activity in Arabidopsis, and they do not explain the distribution of ORI-TEs among the different TE families found here.

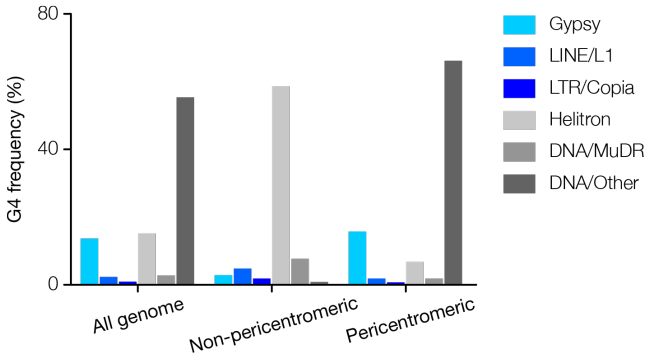


Figure 48 – Colocalization of G4 quadruplexes and TEs. Frequency of the predicted G quadruplexes (G4) in various TE families, as indicated, in whole genome, non-pericentromeric and pericentromeric regions.

ORI-TE activity and the chromatin landscape

We next focused on the chromatin landscape around ORI-TEs to identify a possible common signature. To gain an overall view of the chromatin associated with ORI-containing TEs we looked for possible differences within TE families. We first investigated the whole chromatin signatures associated with the different TE families according to the known Arabidopsis chromatin states (Sequeira-Mendes et al., 2014). Interestingly, we found that the majority of Gypsy, LINE and Copia families, which concentrate more than 80% of all ORI-TEs are associated with chromatin state 9 (Fig. 49A, left panels), which is characteristic of the GC-rich heterochromatin. This is particularly striking for the Gypsy elements, of which ~95% are found in this heterochromatic state. On the contrary, Helitrons, which have a very low tendency to contain ORIs, are not associated with chromatin state 9 but with chromatin states 4 and 8 (Fig. 49A, left panels). Chromatin state 4 is mainly associated with intergenic regions enriched in the Polycomb mark (H3K27me3), whereas chromatin state 8 is an heterochromatin

state characterized by a lower GC content and a higher H3K27me3 level, as compared with the heterochromatin of chromatin state 9 (Sequeira-Mendes et al., 2014). Very interestingly, ORI-containing TEs tend to be in the chromatin state 9, independently of their family (Figure 9A, right panel).

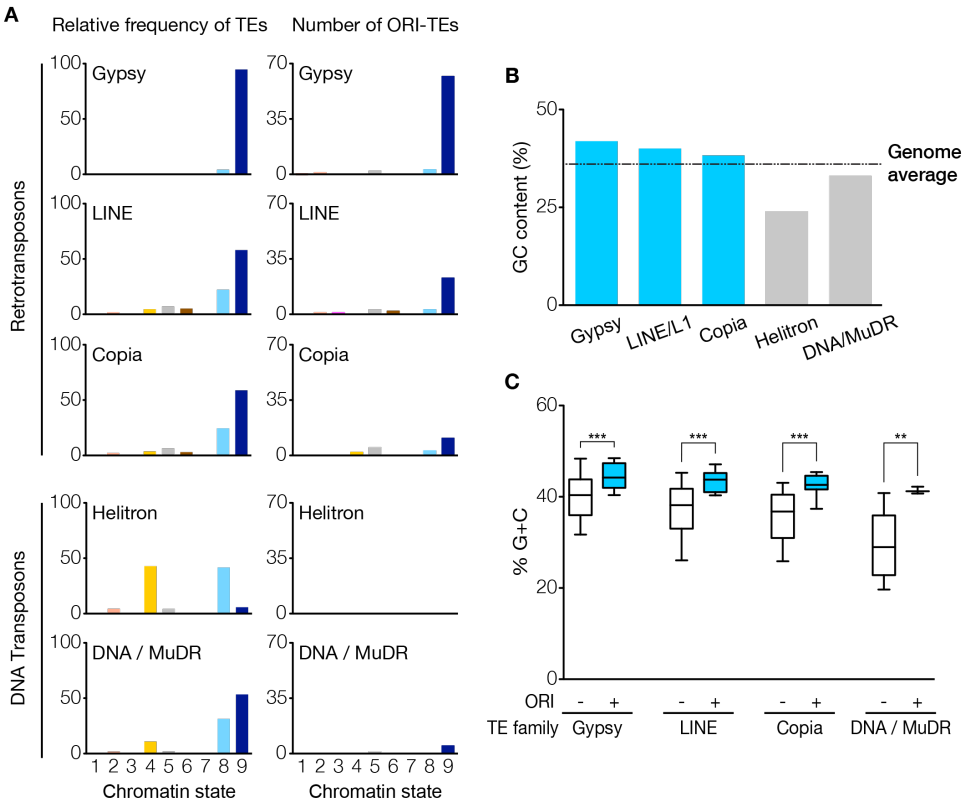


Figure 49 – Chromatin landscape of ORI-TEs and the different TE families. (A) Relative frequency of several TE families (Gypsy, LINE, Copia, Helitron and DNA/MuDR), or ORI-TE of those families with respect to total nucleotide family content, in the nine chromatin states. **(B)** GC content in various Arabidopsis TE families. Retrotransposons (blue) tend to have a higher GC content compared to DNA transposons (grey). **(C)** Average G+C content of TEs with (blue) and without (white) ORIs in the different TE families. ***, p<0.0001; **, p<0.001 (unpaired t-test with Welch's correction; whiskers at 10-90 percentiles, outliers not represented in the graph).

The main feature distinguishing the two heterochromatic states is the GC content, which is higher in chromatin state 9. In fact, this is a striking difference between TEs since the families that tend to contain ORIs (Gypsy, Copia and LINE) have a higher than genome average GC content. For instance, Gypsy elements contain 42.1% GC, the highest among TEs, compared with the 36.5% average GC content of the Arabidopsis genome (Fig. 49B). On the contrary, Helitron elements are characterized by having a very low GC content (24.2%; Fig. 49B). Importantly, however, calculation of the average GC content of TEs that contain ORIs revealed that it was statistically significantly higher than in TEs of the same family that do not contain

ORIs (Fig. 49C). This clearly suggests that a high GC content behaves as a determinant for ORI preference also at heterochromatic loci.

ORI-TE activity is maintained with high H3K9me2 levels

The association of ORI-TEs with a heterochromatin state is somehow surprising as most ORIs are located within genes that colocalize with euchromatic marks found in very different chromatin states (Sequeira-Mendes et al., 2014). Even though we have already shown that transcription of a subset of ORI-TEs is not reactivated in cultured cells (Fig. 46) and because chromatin may undergo changes in some cultured cells (Chupeau et al., 2013), we decided to analyze the chromatin marks associated with ORI-TEs in the Arabidopsis MM2d cultured cells. We first looked at the overall levels of H3K9me2 and H3K27me1, two typical heterochromatic marks that strongly contribute to maintaining the silenced state of TEs in Arabidopsis (Law and Jacobsen, 2010; West et al., 2014), by immunolocalization in cultured cells. H3K27me1 showed a pattern colocalizing with increased DAPI signal whereas H3K9me2 had a dotted appearance in nuclear sites enriched for H3K27me1 and DAPI positive regions (Fig. 50A), as it occurs in the nuclei of Arabidopsis plants. It must be noted that DAPI-stained chromocenters were not very apparent in nuclei of these cultured cells, suggesting a less condensed organization of the pericentromeric heterochromatin compare to seedlings.

To determine more precisely the levels of H3K9me2 and H3K27me1 marks in cultured cells we performed ChIP (using three biological replicates) and analyzed a subset of TEs containing a functional ORI. Although Helitron elements are not associated with ORIs, we also evaluated some Helitron elements located in the two heterochromatin states (AT-rich and GC-rich chromatin states 8 and 9, respectively). In all cases we normalized the measurements to the local H3 content determined by ChIP with anti-H3 antibody. We found that, in all the examples analyzed, the Gypsy and LINE elements (GC-rich heterochromatin state 9) contain a high level of H3K9me2 (Fig. 50B). We also found that in general the H3K9me2 level was higher in retrotransposons than in Helitron elements, independently of their chromatin state (Fig. 50B), similar to what was reported in maize (West et al., 2014). In the case of H3K27me1, which is typical of heterochromatin and crucial to prevent re-replication (Jacob et al., 2010), ChIP experiments revealed that the TEs analyzed showed various levels of H3K27me1 independently of (i) being Gypsy, LINE or Helitron, (ii) their chromatin signature and (iii) their colocalization with ORIs (Fig. 50C).

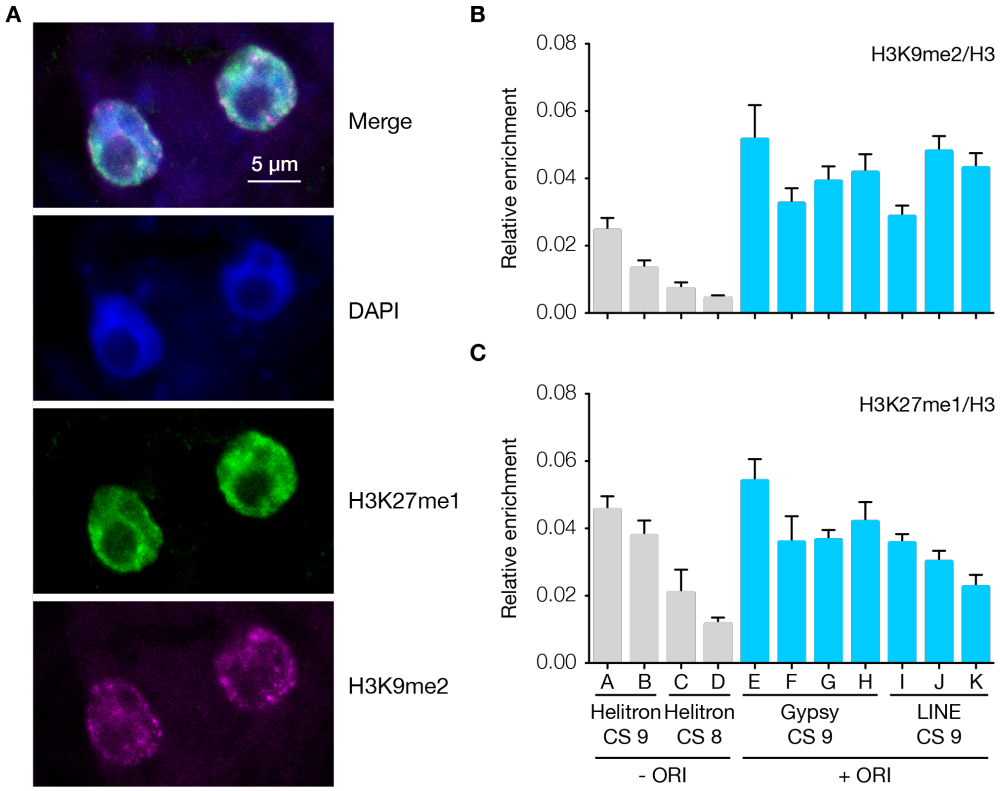


Figure 50 - Heterochromatin marks in Arabidopsis MM2d cultured cells. (A) Immunolocalization of H3K9me2 (magenta) and H3K27me1 (green) in nuclei of cultured cells. Nuclei were stained with DAPI (blue). Levels of H3K9me2 **(B)** and H3K27me1 **(C)** determined by ChIP-qPCR in TEs representative of various families, chromatin states (CS) and with (blue bars) or without ORIs (grey bars). Enrichment values were made relative to the local H3 content determined by ChIP with anti-H3 antibody. Three biological replicates and three technical replicates were evaluated. The mean values ± standard error of the mean is plotted. The codes for the primer pairs used to identify each TE, according to the list in Supplementary Table 4, are: A, AT2TE13970; B, AT4TE16735; C, AT2TE16335; D, AT4TE17050; E, AT4TE16726-2; F, AT4TE16726-3; G, AT1TE62820-3; H, AT1TE62820-5; I, AT2TE15565-2; J, AT2TE15565-3; K, AT4TE03295.

Alterations in the nuclear DNA content are indicative of massive defects in re-replication control and, indirectly, of possible decrease in H3K27me1, as it occurs in the *atxr5/6* mutant (Jacob et al., 2010). Consistent with our ChIP data, we could not detect any significant alteration in the DNA content profile of cultured Arabidopsis cells (Fig. 51). Since retrotransposons are enriched for ORIs and H3K9me2 although there is a lack of correlation of H3K27me1 with ORI-TEs, the lack of this mark in the Arabidopsis cultured cells is unrelated to ORI-TE activity at heterochromatin.

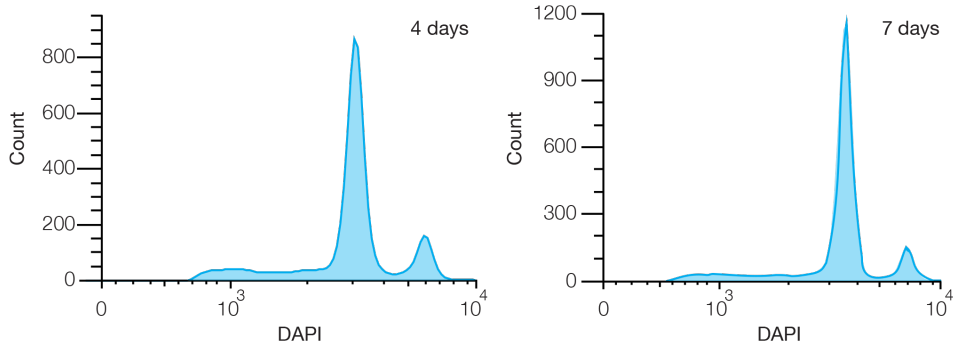


Figure 51 – Flow cytometry profile of Arabidopsis MM2d cells used in this study. The DNA content distribution of DAPI-stained nuclei of cultured Arabidopsis cells was estimated by flow cytometry at the indicated times after subculturing.

5. Discussion

5.1 *ORC1a* and *ORC1b* have acquired different functions during plant evolution

Several full or partial genome duplication events have taken place during plant evolution. As a consequence, many genes are duplicated and evolve independently. *ORC1*, located in a duplicated region of chromosome 4, is one of those examples (AGI, 2000). A recent study differentiates *ORC1a* and *ORC1b* as newly duplicated genes in *Arabidopsis thaliana* (Wang et al., 2013). Previous analysis (Diaz-Trivino et al., 2005), revealed the different expression domains of these two genes. Using translational fusion constructs we have demonstrated that the ORC1a and ORC1b proteins have distinct expression domains in young seedlings. ORC1a is clearly associated with endocycling cells while ORC1b is present in both endocycling and proliferating cells. Moreover, the patterns of expression within the nucleus of the Arabidopsis root cells were quite different (Fig. 15, 52). While ORC1a associates with chromocenters almost exclusively, ORC1b colocalizes with both euchromatic and heterochromatic regions. ORC1b protein expression is different in proliferating and endocycling cells. The highest levels of the protein are detected in cycling cells. In endocycling cells two main changes related to ORC1b content are observed: a reduction of the protein levels and a more avid association with chromocenters.

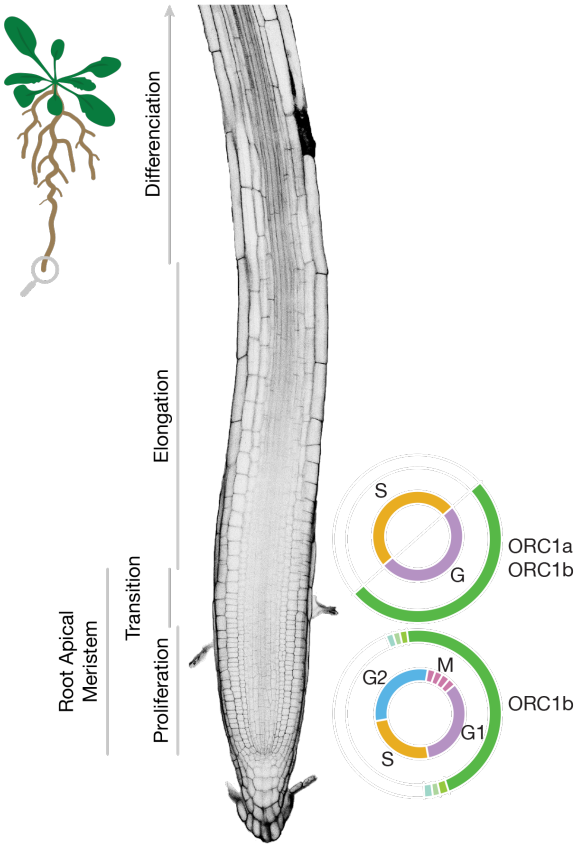


Figure 52 – Dynamics of ORC1 proteins during cell cycle and endocycle at the root meristem. ORC1b is present in proliferating cells from G2 to the G1/S transition. In the endocycling cells ORC1a and ORC1b proteins are only present during the G-phase.

So far, it is unknown whether the mechanisms of licensing, ORI selection and activation are maintained between cell cycle and endocycle. However, there is evidence suggesting that different mechanisms may occur. In Drosophila, ORC1 is dispensable for endoreplication, while other pre-RC components such as CDT1 or MCM are essential (Park and Asano, 2008). Very recently, a stable mutant of ORC1 in human cells was obtained (Shibata et al., 2016). In the absence of ORC1, cells could grow correctly, but CDC6 became essential for DNA replication. Remarkably, ORC1 and CDC6 have a common evolutionary origin, and in Archeas and protists a single ORC1/CDC6 protein is present (Ausiannikava and Allers, 2017; Godoy et al., 2009; Kumar et al., 2008). During endoreplication in Arabidopsis both ORC1a and ORC1b are present, although both proteins are detected at low amounts and associated to heterochromatin. If the role of ORC1a and ORC1b were to license the ORIs during endocycle, theoretically, an increase in the amount of protein would resemble the increase in DNA content in every endocycle, which is not the case. Previous studies in Arabidopsis revealed that *CDC6* was necessary for the endocycle and that the *CDC6* gene was overexpressed in hypocotyls grown in the dark (Castellano et al., 2001). All these data lead us to hypothesize that different mechanisms of licensing could operate during cell cycle and endocycle. Thus, in proliferating cells ORC1b would drive licensing of ORIs and during endocycle CDC6 could have a more important role in this process. Still, we observed that during the endocycle both ORC1a and ORC1b proteins are present at the G-phase, hence, we cannot rule out the possibility that the two proteins play a role in ORI licensing during endocycle. However, in that case, it would be restricted to the heterochromatic sites were the proteins are present. We also found that *orc1b-1* mutant plants present a regular growth in normal conditions. It is possible that in this situation other protein of the pre-RC complex, for instance CDC6, would lead the licensing and for that reason the plants are viable, similar to what happens in ORC1 mutant human cells (Shibata et al., 2016). Nevertheless, “CDC6-licensing” during G1 would not be as effective as the canonical ORC1-mediated pathway, because in the presence of aphidicolin *orc1b-1* mutant root growth is severely impaired (see below).

The overall dynamics of the ORC1b protein of Arabidopsis (Fig. 52) resembles the dynamics of the ORC1 in humans. First, both proteins are synthesized at low levels in G2 and remain bound to chromatin during the entire mitosis. Second, the levels of the protein increase considerably during ORI licensing in G1. Finally, the two proteins are degraded at the G1/S transition by the ubiquitin proteasome pathway. Importantly, the degradation of ORC1b take place at the G1/S transition, which implicates that the rescue of dormant ORIs or the firing of late replicating ORIs

would be independent of ORC1b. However, both dormant and late ORIs would be already licensed when ORC1b is degraded, indicating that the degradation of ORC1b is a mechanism to avoid re-replication.

Remarkably, other proteins of the pre-RC complex such as ORC2, ORC6, CDT1a and MCM7 present different expression patterns in the Arabidopsis root, compared to ORC1. While ORC6 (Diaz-Trivino, 2005) and MCM7 (Herridge et al., 2014) are present during the entire cell cycle except in mitosis. In other organisms where the ORC complex is not removed from the chromatin, for instance in yeasts, a new ORC hexamer, in the inactive form, binds to the ORI site during S-phase, in order to maintain the ORI sites. An analogous mechanism could be occurring in Arabidopsis cells, not during S-phase, but instead, during G2. At the G1/S transition ORC1b is rapidly degraded, and the rest of the complex (ORC2 and ORC6) remains bound to the chromatin. After DNA replication ORC1b is synthesized and the new protein could bind to the sites were ORC is already present.

ORC1a and ORC1b proteins present distinct regulation, whereas the amino acid sequence is similar in 88% of the residues. While ORC1b show a clear accumulation upon inhibition of the ubiquitin proteasome pathway, ORC1a appears delocalized from the chromocenters without an increase in the signal. The observation for ORC1a suggests that a monoubiquitination process could be leading to a delocalization of the protein. It has been reported in mouse embryonic thymus (Miyake et al., 2005) a splicing variant of ORC1 (called Orc1A and the variant Orc1B). Orc1B variant was smaller, and localized at the cytoplasm. Interestingly, the degradation of Orc1B occurs in a proteasome independent manner. This resembles the situation in Arabidopsis, where a second protein (in this case is a new duplicated gene), acquires a distinct degradation mechanism. ORC1a-GFP levels were maintained normal in the mutant background for the F-box proteins *skp2a-1* and *fbl17-1.* On the contrary, ORC1b-GFP was highly accumulated in the mutant background *fbl17-1* and remains normal in the mutant background of *skp2a-1*. In humans, the SCF^{Skp2} E3-ubiquitin ligase recognizes ORC1 triggering its degradation (Mendez et al., 2002). This result suggests that the functional homologue for the F-box Skp2 could be FBL17, instead of the amino acid sequence homologue SKP2A. In humans, ORC1 is phosphorylated prior to polyubiquitination and degradation. In Arabidopsis this does not seem to be the case. After treatment with roscovitin, which specifically inhibits CDK phosphorylation, ORC1b levels were not increased but decreased. ORC1a and ORC1b amino acid sequence is identical in 88% of the residues. The N-terminal part of the proteins contains the different amino acids (Fig. 53). In that

region, specific CDK phosphorylation sites and SUMOylation sites are predicted for both ORC1a and ORC1b proteins.

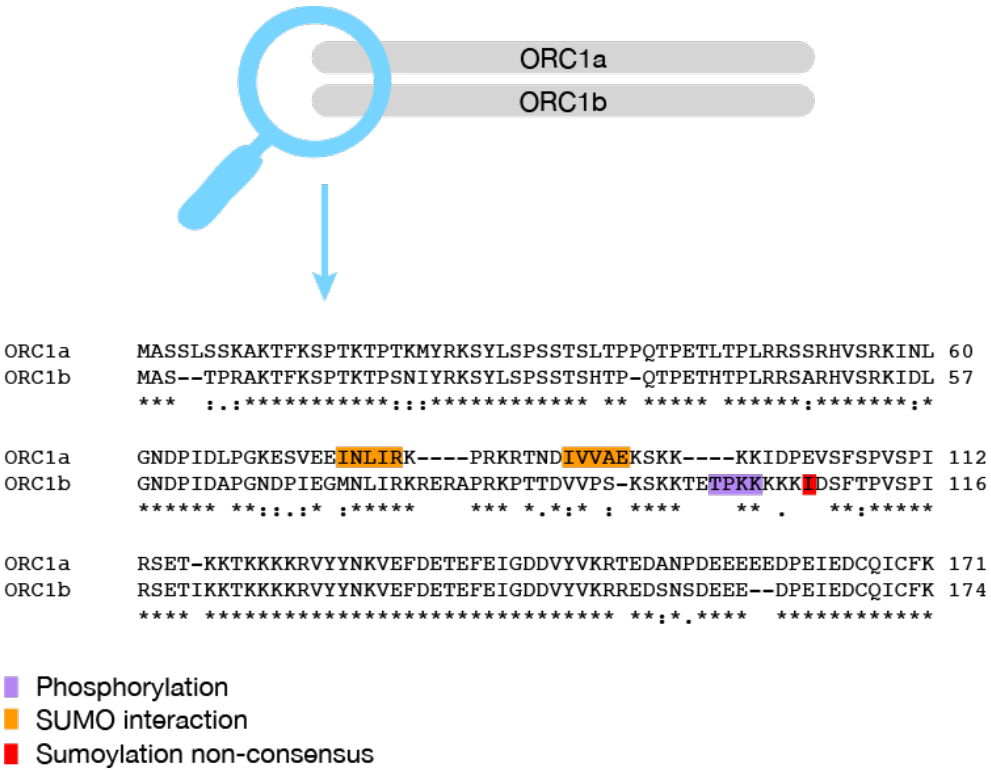


Figure 53 – N-terminal region of Arabidopsis ORC1a and ORC1b proteins. Amino acid residues were aligned and the region containing most of the differences between ORC1a and ORC1b is shown. CDK phosphorylation, SUMO interaction and sumoylation non-consensus sites exclusive for either ORC1a or ORC1b are shown.

The signal that would trigger the recognition of Arabidopsis ORC1b by SCF^{FBL17} for degradation would be localized to this N-terminal region. Three situations may explain the observed outcome (i) another kinase, insensitive to roscovitin, phosphorylates ORC1b at the specific site; (ii) a different modification of ORC1b, produced by a different interacting partner could trigger degradation, for instance, human Orc1 interacts with the histone acetylase HBO1 (Iizuka and Stillman, 1999); (iii) instead of a modification a conformational change due to the binding or delocalization of the pre-RC proteins during ORI activation could trigger the recognition by the E3-ubiquitin ligase.

Strikingly, the two *orc1* mutant plants exhibit a normal growth under standard conditions even though they present a delay in S-progression, specifically in the progression of early S-phase. In humans, mutations in the pre-RC complex and especially in the *ORC1* gene lead to Meier-Gorlin syndrome (MGS), immunodeficiency and other rare diseases (Munoz and Mendez, 2017).

Accordingly, mutations in the bromo-adjacent homology (BAH) domain of ORC1 in *X. laevis* cause a dwarfish phenotype, similar to MGS (Kuo et al., 2012). The uncoupling between the problems in S-phase and the absence of growth defects in *orc1* mutants, and especially in *orc1b-1* mutants, suggest that ORI licensing is partially affected to levels compatible with normal growth.

While *orc1a-1* survives to DNA replication stress situations like the wild type plant, *orc1b-1* mutants are hypersensitive to aphidicolin (Fig. 54). Also, *orc1b-1* plants tolerate zeocin and hydroxyurea (HU) like the wild type plant.

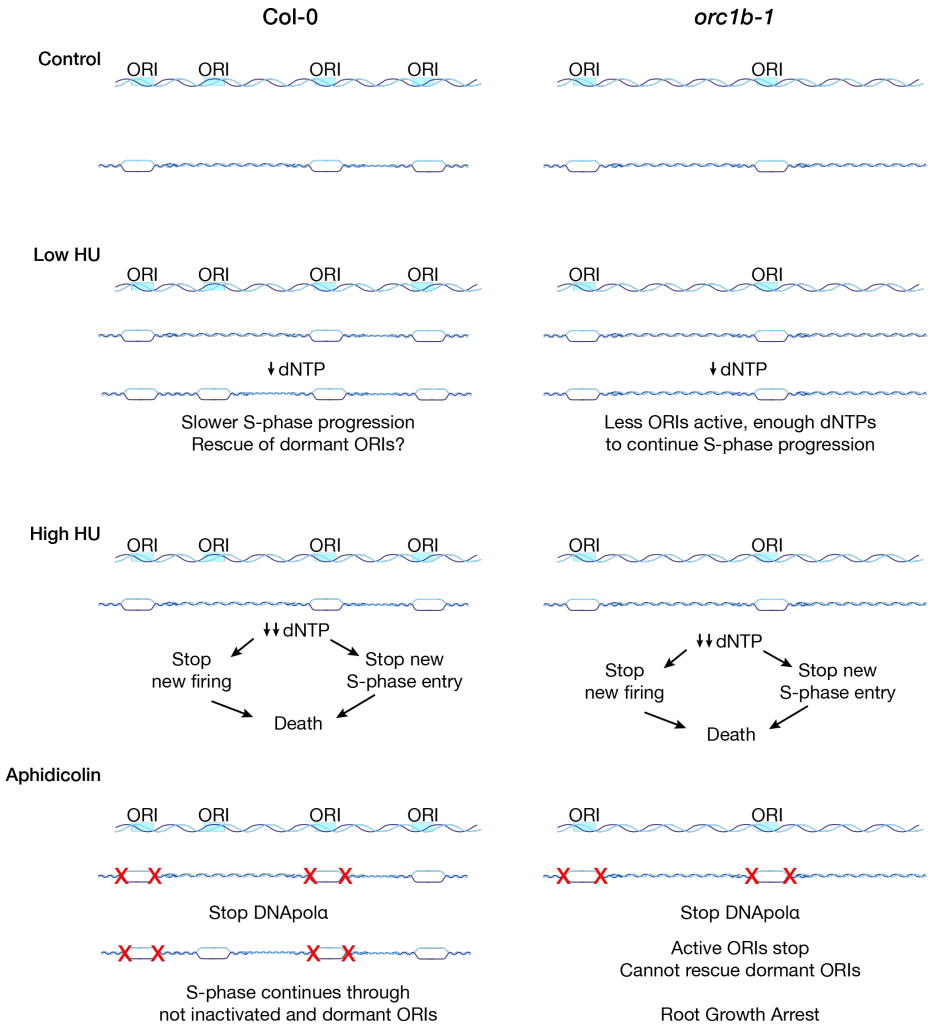


Figure 54 – Model for the rescue of dormant ORIs during HU and aphidicolin stresses. In Col-0 wild type plants, low concentration of HU produce a slower S-phase and probably rescue of dormant ORIs. *orc1b-1* mutant plants, which have less active ORIs can overcome low concentration of HU like the wild type plants by continue with a slow S-phase. Upon high HU concentration both wild type and orc1b-1 mutant plants die. Aphidicolin stops the forks by inhibition of the DNA polymerase α (DNApolα). In the wild type plants, with high density of functioning ORIs, some will not be affected by aphidicolin and continue the progression. S-phase will be completed by the rescue of dormant ORIs. In orc1b-1 mutant plants, with fewer ORIs active, aphidicolin will block the large majority of them, and after the impossibility of rescuing dormant ORIs, the root growth stops.

Hydroxyurea and aphidicolin produce similar effects, stalling of the progressing forks, but through two different mechanisms. HU inhibits the ribonucleotide diphosphate reductase, causing a depletion in the deoxynucleotide pool, while aphidicolin targets primarily the DNA polymerase α. In wild type plants, both stresses are overcome by rescuing dormant ORIs. On the contrary, *orc1b-1* mutants, which already have less licensed ORIs, can only survive to the HU stress. *orc1b-1* already have a longer S-phase, so in the presence of HU, instead of rescuing dormant ORIs, they may continue with a delayed S-phase until completing it entirely. However, when the stress is produced by aphidicolin treatment, the forks will be completely stopped, and the only way to complete DNA replication is through the rescue of dormant ORIs, which *orc1b-1* fails to do because a reduced ORI licensing. In plants there are a few examples (De Schutter et al., 2007) where HU and aphidicolin stresses produce different outcomes.

A very different phenotype is present in *orc1a-1* mutants. Plants lacking ORC1a have less H3K27me1 at the chromocenters of some endocycling cells. In other eukaryotic systems, ORC1 binds to heterochromatin related proteins such as Sir1 in *S. cerevisiae* (Hou et al., 2005) or HP1 in Drosophila and humans (Prasanth et al., 2010; Shareef et al., 2001). The H3K27me1 mark is specific of plant heterochromatin and is deposited by the methyltransferases ATXR5 and ATXR6 (Jacob et al., 2009). ATXR5/6, binds to PCNA at the forks to deposit the mark in the newly synthesized DNA. One attractive possibility is that during the G phase of the endocycle, ORC1a recruits ATXR5/6 to the ORIs of the heterochromatin regions (Fig. 55). This could help methyltransferases to be already in the ORI site prior to the ORI firing. Then, in S-phase, ATXR5/6 could be transferred from the pre-RC to the replisome to deposit the H3K27me1. In the absence of ORC1a, the recruitment of ATXR5/6 would be impaired, causing the loss of the mark. Additionally, TREX-2 complex was identified in a screening for suppressor mutations of the *atxr5/6* mutant (Hale et al., 2016). TREX-2 absence alleviates the DNA replication stress of the *atxr5/6* mutants, which is caused by the massive expression of TEs during S-phase. Interestingly, the Drosophila ORC complex interacts with TREX-2 (Kopytova et al., 2016). Therefore, a third possibility is that ORC1a interacts with ATXR5/6 in a TREX-2 dependent manner, facilitating the recruitment of ATXR5/6 to the ORIs.

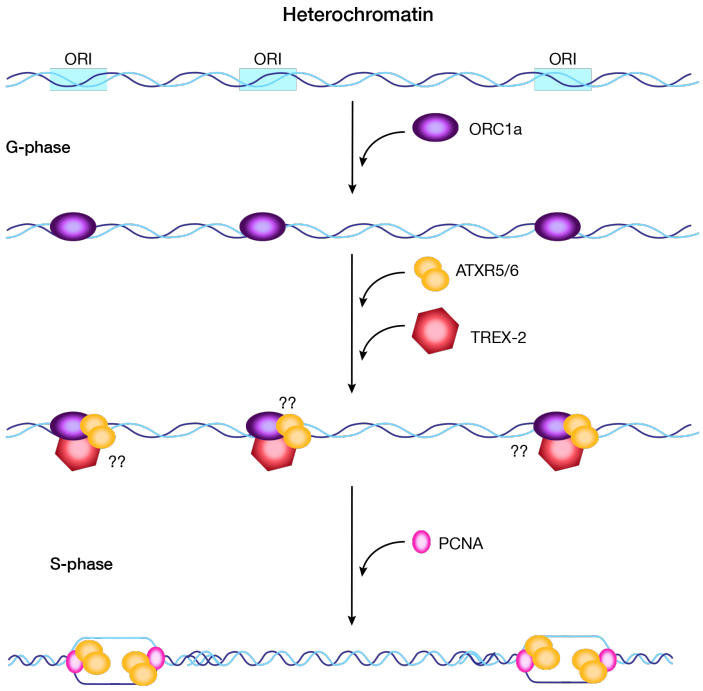


Figure 55 – Model of the pathway ORC1a – ATXR5/6 – TREX-2. ORC1a is present during G-phase at the heterochromatin ORIs to recruit ATXR5/6 in a TREX-2 dependent manner. Upon S-phase entry, ATXR5/6 is already at the ORI sites, interacts with PCNA at the open replication bubbles to maintain H3K27me1 levels.

Taking all the results together, we propose that ORC1a and ORC1b proteins have acquired different functions in plants, while in animals the same protein may play a dual role, during ORI licensing and preserving heterochromatin structures. Remarkably, the relationship between ORC1 and the mechanism of chromatin silencing is ancient in evolution. In yeasts, the *ORC1* gene was also duplicated, and one of the copies gave rise to the current *Sir3* gene of *S. cerevisiae* (Hickman and Rusche, 2010). In the yeast *Kluyveromyces lactis*, which diverged from *S. cerevisiae* prior to the duplication of the *ORC1* gene, ORC1 interacts with Sir2 and Sir4 to establish heterochromatin. The ORC1 from Drosophila, Xenopus and mammals interacts with HP1 (Prasanth et al., 2010; Shareef et al., 2001) during the establishment and maintenance of the heterochromatin. As newly duplicated genes in Arabidopsis (Wang et al., 2013), *ORC1b*, the parental gene, retains the canonical function driving ORI licensing. On the contrary, the new gene *ORC1a* presents a specialized function in a group of cells during endocycle. It is possible that the evolutionary history that occurred in the yeast lineage is repeated in the plant lineage with ORC1a and ORC1b, while in animals with a unique ORC1 protein, the two functions are carried out by the same protein.

5.2 When replication meets heterochromatin

Over the past years, detailed genome-wide maps of ORIs have been generated for various multicellular organisms such as cultured Drosophila, mammalian and Arabidopsis cells (Cayrou et al., 2015; Comoglio et al., 2015; Costas et al., 2011). ORI specification and activation depends on several variables, including the cell type and the physiological state as well as specific chromatin features, frequently including those associated with open chromatin (Sequeira-Mendes and Gutierrez, 2015). The common observations of ORIs across multicellular organisms studied are that they (i) preferentially colocalize with genic regions, in particular with highly expressed genes, (ii) correlate with high GC content, and (iii) tend to be present in sequences that may form DNA tertiary structures such as G-quadruplexes (G4). However these ORI specification studies focused on how euchromatic ORIs are selected and the specific characteristics that define them. Some previous studies reported the late replication timing of heterochromatin both in animals and plants (Lee et al., 2010), the peculiar ORI features in the facultative inactive chromosome X or the links between DNA replication fork progression and the establishment of heterochromatin (Gutierrez et al., 2016; Nikolov and Taddei, 2016). Nonetheless, the genomic features that contribute to specify ORIs in constitutive heterochromatin have not been studied and, consequently, are poorly understood. In general the constitutive heterochromatic sites, located at the centromeres, telomeres or the nucleolus organizing region, replicate late. This can be easily imaged by immunocytochemical assays detecting thymidine analogues. In Arabidopsis, it was described that chromosome 4 replicates in two phases, early and late, colocalizing the last one with the vast majority of heterochromatin (Lee et al., 2010), but no further details related to ORI specification at this sites were analyzed.

The identification of ORIs responsible for replication of pericentromeric heterochromatin requires reliable genome annotation, particularly for the repeated sequences that conform heterochromatin. The Arabidopsis genome was sequenced >15 years ago (AGI, 2000), and an updated genome annotation (TAIR10), including highly repetitive pericentromeric regions, is now available, making possible the generation of a high-resolution map of ORIs, paying particular attention to those located in heterochromatic regions. Using TAIR10 annotation, we found that Arabidopsis ORIs in euchromatin regions are almost exclusively located within genes while in the heterochromatic pericentromeric regions a significant fraction of ORIs colocalize with TEs. This suggests that in the absence of genes, TEs are selected to act as DNA replication origins. Since pericentromeric heterochromatin is packed with repeated sequences, and we based our mapping

analysis in unique reads, it is likely that the actual number of ORI-TEs is larger.

TEs constitute a very heterogeneous type of repetitive elements that can be divided in different classes and families (Deragon et al., 2008; Wicker et al., 2007). Our study also shows that not all TEs serve equally as ORIs. Retrotransposons, and in particular Gypsy elements, more frequently colocalize with them, while ORI-TEs are excluded from the Helitron DNA transposon family. The striking tendency of ORIs to colocalize with retrotransposons is maintained even at chromosome arms, where DNA transposons are more frequent. The second most prevalent family of ORI-TEs is the LINE retrotransposons. This opens an important question about the possible role of LINE elements in mammalian DNA replication origins, where they are the most prevalent family of TEs (Kazazian, 2000).

The activity of ORIs has been frequently associated with the expression level of the genomic loci where they are located. In fact, in some systems the specific cell-type replication program mimics the specific cell-type transcription program (Comoglio et al., 2015; Rodriguez-Martinez et al., 2017). Two scenarios, related to DNA replication, have been described in which there was no associated transcription: during cleavage cell divisions of *X. laevies* and *D. melanogaster* embryos, and in the inactive chromosome X of mammalian cells. In the first case, DNA replication initiates from randomly distributed loci (Hyrien and Mechali, 1993; Sasaki et al., 1999), and in the latter the replication pattern resembles a random process (Koren and McCarroll, 2014). This suggests that in the absence of transcription (the inactive chromosome X is a special type of facultative heterochromatin), DNA replication may initiate randomly. However, we found no association between TE expression and ORI-TE activity, suggesting that this particular type of ORIs located at the constitutive heterochromatin is independent of the transcription status throughout the ORI region. Nonetheless, transcription is the first and obligate step for mobilization of all retrotransposons, whereas DNA transposons are mobilized by a DNA intermediate and do not need to be transcribed. This makes retrotransposons more similar to genes than any other TE. Indeed, whereas the majority of retrotransposons are silent in most plant tissues, their activation under stress or in particular mutant backgrounds confirms that they retain the capacity to be transcribed and to transpose (Bucher et al., 2012; Cavrak et al., 2014). Another condition known to produce TE reactivation is immortalization of cells in culture. Although the cells lines used by Tanurdzic and colleagues (Tanurdzic et al., 2008) and by us are different, we found that the vast majority of ORI-TEs were not located in TEs reactivated in their cell culture. Therefore, our results show that the activity of ORI-TEs cannot be explained by transcription through TE sequences.

The central region around the ORIs in euchromatin tends to be depleted of CG methylation (Costas et al., 2011). We found that C methylation for the three sequence contexts was lower in Helitron elements, which do not contain any ORI, while Gypsy elements were highly methylated, according to previous studies (Ahmed et al., 2011). Furthermore, Gypsy C methylation levels were high independently of the colocalization with ORIs, suggesting that C methylation does not affect ORI selection. This is also the case for the inactive chromosome X, where C methylation levels do not affect ORI usage (Gomez and Brockdorff, 2004).

DNA tertiary structures such as G4 have been frequently related to ORI selection in multicellular eukaryotes (Cayrou et al., 2015). In particular, it has been suggested that the ORC complex could specifically recognize structures during the licensing process (Hoshina et al., 2013). However, we found that in the Arabidopsis genome, G4 motifs were more frequent in TEs than in genes, although ORIs tend to colocalize with genes. Moreover, most of the G4 motifs accumulate in ATREP18 elements, a TE family highly depleted in ORIs. Thus, our results do not support the idea that G4 structures explain the distribution of ORIs (whether in genes or in TEs). The peculiarities of the plant ORC complex, and specially the structure of the ORC1 subunit, open the possibility of a plant-specific mechanism of recognition, maybe involving other tertiary structures different from the G4.

ORIs located within genes in the euchromatin are mostly associated with open chromatin, although they can be found in very different chromatin states (Sequeira-Mendes and Gutierrez, 2015). In Arabidopsis, there are two types of heterochromatin states. Both are highly enriched in the typical plant heterochromatin marks (C methylation, H3K9me2 and H3K27me1; (Sequeira-Mendes et al., 2014). We found that ORI-TE activity occurs while maintaining high levels of H3K9me2 and H3K27me1, indicating that ORI selection and activation is allowed even at closed heterochromatic sites. Nevertheless, how the ORC complex recognizes and binds to these sites is still unclear. In eukaryotes, the ORC complex interacts with several proteins important for the heterochromatin formation, such as HP1 from Drosophila, Xenopus and human cells (Pak et al., 1997; Prasanth et al., 2010) or Sir1 in *S. cerevisiae* (Triolo and Sternglanz, 1996). This suggests that structural heterochromatin proteins may mediate the recognition of the heterochromatin ORIs by ORC. Independent of the mechanism, plants expressing ORC1 (either ORC1a or ORC1b proteins) fused to a GFP tag, show high affinity for chromocenters and other heterochromatic structures.

The two heterochromatin states found in Arabidopsis (Sequeira-Mendes et al., 2014) are discriminated depending on the GC content: chromatin state 8 is AT-rich while chromatin state 9 is

GC-rich. Both genes and retrotransposons show an above average GC content in Arabidopsis, which makes their sequences different from most DNA transposons and particularly Helitron elements. Importantly, TEs with ORIs possess a higher GC content than TEs without ORIs, independently of their TE family. Therefore, these results lead us to propose that a high local GC content, typical of the heterochromatin state 9 where the vast majority of ORI-TEs are located, together with the potential to be transcribed, characteristic of the genomic organization of retrotransposons, are the major features of ORIs colocalizing with TEs. These characteristics allow certain TE families to contribute to a significant fraction of ORIs in heterochromatic regions. Whereas the prevalence of Gypsy retrotransposons is particular to plants, retrotransposons make up an important fraction of the genome, and in particular of heterochromatic regions, of both plants and animals. For example, in mammals the L1 LINE accounts for as much as 20% of the genome and is enriched in heterochromatin (Graham and Boissinot, 2006). Although LINEs represent a small fraction of Arabidopsis pericentromeric regions, we show that they can significantly contribute to ORI-TE specification. Although mouse ORIs have been mapped extensively in relation to their chromatin context, a detailed localization of late-replicating ORIs, likely corresponding to heterochromatic regions, to specific genomic elements has not been undertaken (Cayrou et al., 2015). Hence, it is tempting to speculate that the contribution of retrotransposons to ORI specification in heterochromatin shown here for Arabidopsis could also be important in other eukaryotic species.

When TEs were first described by Barbara McClintock (McClintock, 1944), they were considered “junk DNA” and “parasitic elements”. For the last ten years, more evidence is accumulated suggesting that TEs are an important evolutionary force in shaping the eukaryotic genomes (Chenais et al., 2012; Mita and Boeke, 2016). First, TEs give mechanical support to the centromeres to be compact. This is highly important for chromosome segregation during mitosis and meiosis. Mutants in the enzymes that deposit specific heterochromatin marks present loose chromocenters (for instance, *fas1-4*, *ddm1* or *atxr5/6* (Exner et al., 2006; Jacob et al., 2009; Zemach et al., 2013)) and have problems during mitosis such as formation of mitotic bridges, which utterly lead to genomic instability. Second, TEs are a source of variability. In the presence of environmental stress conditions, TEs expression is reactivated, leading in some cases to new transposition events (Cavrak et al., 2014; Negi et al., 2016; Xu et al., 2015). TEs contain binding sites for multiple TFs (Henaff et al., 2014) and upon transposition these TF binding sites will regulate the expression of genes nearby the transposition site (Chuong et al., 2017). In some

cases, new transpositions events produce new proteins due to the appearance of new splicing sites (Chuong et al., 2017). And finally, we demonstrate that TEs participate in fundamental genomic processes such as DNA replication. Retrotransposons, more similar to genes than other type of TE, provide the structures necessary for ORI initiation at pericentromeric gene-poor regions. Otherwise, ORIs would be forced to colocalize with intergenic regions, which in Arabidopsis are AT-rich. On the contrary, retrotransposons are GC-rich and in multicellular eukaryotes high GC content is a common denominator for ORIs selection, in both euchromatin and heterochromatin region.

6. Conclusions

1. *ORC1a* and *ORC1b* genes exhibit different expression domains during root organogenesis. While ORC1b is present in both proliferating and endocycling cells, ORC1a is restricted to endocycling cells and preferentially associates with heterochromatin. ORC1b expression patterns change along the root meristem, the highest levels are detected in the proliferation zone where the protein associates with euchromatin and heterochromatin. However, during endocycle, protein levels are lower and mostly associate with heterochromatin.

2. ORC1b is synthesized in G2 and remains bound to the chromatin during the entire mitosis. ORC1b continues loading since telophase and during G1, reaching a maximum 1.5 h after mitosis. At the G1/S transition the E3-ubiquitin ligase SCF^{FBL17} recognizes ORC1b and the protein is degraded in less than 10 min. On the contrary, ORC1a is not a target of the proteasome degradation.

3. Both ORC1a and ORC1b are present during the G-phase of the endocycle, absent during the S-phase and preferentially associated with heterochromatin.

4. The absence of ORC1 in loss of function (knock-out) *orc1a-1* and *orc1b-1* mutant plants does not produce a detectable macroscopic phenotype, including the root meristem, under standard growth conditions.

5. The two ORC1 proteins exhibit different functions, although they are 88% identical in their amino acid sequences. *orc1b-1* mutant plants present a delay in S-phase progression, particularly in early and mid S-phase and are hypersensitive to aphidicolin suggesting that they have less licensed ORIs. On the contrary, ORC1a participates in the establishment of H3K27me1 at chromocenters and the mark is severely diminished in nuclei undergoing endocycle of *orc1a-1* mutant plants.

6. Arabidopsis ORIs preferentially colocalize with genes. At the gene-poor pericentromeric regions, a large fraction of ORIs colocalize with retrotransposons, and in particular of the Gypsy family.

7. ORI-TE activity is independent of TE transcription and G4 structures. ORI-TEs are active while maintaining high levels of C methylation, H3K9me2 and H3K27me1 heterochromatin marks.

8. ORI-TEs colocalize with a specific chromatin signature defined by GC-rich heterochromatin. TEs with active ORIs contain a local GC content higher than the TEs lacking them. ORI colocalization with retrotransposons seems to be determined by their transposition mechanism, based on transcription, and a specific chromatin landscape.

7. Conclusiones

1. Los genes *ORC1a* y *ORC1b* muestran un dominio de expresión diferente durante la organogénesis de la raíz. Mientras que ORC1b está presente en las células tanto proliferantes como en endociclo, ORC1a se restringe a las células en endociclo y se asocia preferiblemente a heterocromatina. El patrón de expresión de ORC1b cambia a lo largo del meristemo radicular, detectándose los mayores niveles en la zona de proliferación donde la proteína se asocia a eucromatina y heterocromatina. En cambio, durante el endociclo los niveles de proteína bajan y se encuentra sólo en heterocromatina.

2. ORC1b se sintetiza en G2, permanece unida a la cromatina durante toda la mitosis y se continúa cargando desde telofase y durante G1, alcanzando un máximo 1.5 h después de la mitosis. En la transición G1/S, la E3 ubiquitina ligasa SCF^{FBL17} reconoce ORC1b degradando a la proteín en menos de 10 minutos. Por el contrario, ORC1a no es una diana de la degradación por vía proteasoma.

3. Ambas proteínas ORC1a y ORC1b están presentes durante la fase G del endociclo, ausentes durante la fase S y preferiblemente asociadas con la heterocromatina.

4. La ausencia de ORC1 en los mutantes de pérdida de función *orc1a-1* y *orc1b-1* no produce ningún fenotipo macroscópico detectable, incluyendo el meristemo de la raíz en condiciones de crecimiento estándar.

5. Las dos proteínas ORC1 presentan funciones diferentes, si bien su secuencia es idéntica en un 88% de los residuos. Las plantas mutantes *orc1b-1* presentan un retraso en la progresión de la fase S, especialmente entre las fases inicial e intermedia. Además, son hipersensibles a afidicolina, lo que sugiere que tienen menos ORIs licenciados. Por el contrario, ORC1a participa en el establecimiento de la marca H3K27me1 en los cromocentros y la marca está severamente disminuida en los núcleos del endociclo de las plantas mutantes *orc1a-1*.

6. Los ORIs de Arabidopsis se localizan preferentemente en los genes. En las zonas pericentroméricas, pobres en genes, una gran parte de los ORIs se localizan en retrotransposones, en especial de la familia Gypsy.

7. La actividad ORI-TE es independiente de la transcripción del TE y de estructuras G4. Los ORI-TEs están activos mientras mantienen niveles elevados de las marcas de heterocromatina metilación en C, H3K9me2 y H3K27me1.

8. Los ORI-TEs colocalizan con una cromatina característica, definida por ser rica en GC. Los TEs con ORIs activos tienen un mayor contenido local de GC que los TEs que carecen de ORIs. La colocalización de los ORIs con los retrotransposones se determina por el mecanismo de transposición, basado en la transcripción y por el paísaje cromosómico específico.

8. References

AGI (2000). Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature *408*, 796-815.

Ahmed, I., Sarazin, A., Bowler, C., Colot, V., and Quesneville, H. (2011). Genome-wide evidence for local DNA methylation spreading from small RNA-targeted sequences in Arabidopsis. Nucleic Acids Res *39*, 6919-6931.

Aida, M., Beis, D., Heidstra, R., Willemsen, V., Blilou, I., Galinha, C., Nussaume, L., Noh, Y.S., Amasino, R., and Scheres, B. (2004). The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. Cell *119*, 109-120.

Araki, M., Yu, H., and Asano, M. (2005). A novel motif governs APC-dependent degradation of Drosophila ORC1 in vivo. Genes Dev *19*, 2458-2465.

Ausiannikava, D., and Allers, T. (2017). Diversity of DNA Replication in the Archaea. Genes (Basel) *8*.

Barrada, A., Montane, M.H., Robaglia, C., and Menand, B. (2015). Spatial Regulation of Root Growth: Placing the Plant TOR Pathway in a Developmental Perspective. Int J Mol Sci *16*, 19671-19697.

Barry, E.R., and Bell, S.D. (2006). DNA replication in the archaea. Microbiol Mol Biol Rev *70*, 876-887.

Barutcu, A.R., Lian, J.B., Stein, J.L., Stein, G.S., and Imbalzano, A.N. (2017). The connection between BRG1, CTCF and topoisomerases at TAD boundaries. Nucleus, 1-6.

Bell, S.P., and Labib, K. (2016). Chromosome Duplication in Saccharomyces cerevisiae. Genetics *203*, 1027-1067.

Bell, S.P., Mitchell, J., Leber, J., Kobayashi, R., and Stillman, B. (1995). The multidomain structure of Orc1p reveals similarity to regulators of DNA replication and transcriptional silencing. Cell *83*, 563-568.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128-134.

Benfey, P.N., Linstead, P.J., Roberts, K., Schiefelbein, J.W., Hauser, M.T., and Aeschbacher, R.A. (1993). Root development in Arabidopsis: four mutants with dramatically altered root morphogenesis. Development *119*, 57-70.

Bennetzen, J.L., and Wang, H. (2014). The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol *65*, 505-530.

Berckmans, B., and De Veylder, L. (2009). Transcriptional control of the cell cycle. Curr Opin Plant Biol *12*, 599-605.

Bernatavichute, Y.V., Zhang, X., Cokus, S., Pellegrini, M., and Jacobsen, S.E. (2008). Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in Arabidopsis thaliana. PLoS One *3*, e3156.

Besnard, E., Babled, A., Lapasset, L., Milhavet, O., Parrinello, H., Dantec, C., Marin, J.M., and Lemaitre, J.M. (2012). Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol *19*, 837-844.

Bicknell, L.S., Walker, S., Klingseisen, A., Stiff, T., Leitch, A., Kerzendorfer, C., Martin, C.A., Yeyati, P., Al Sanna, N., Bober, M.*, et al.* (2011). Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. Nat Genet *43*, 350-355.

Bleichert, F., Botchan, M.R., and Berger, J.M. (2015). Crystal structure of the eukaryotic origin recognition complex. Nature *519*, 321-326.

Blilou, I., Xu, J., Wildwater, M., Willemsen, V., Paponov, I., Friml, J., Heidstra, R., Aida, M., Palme, K., and Scheres, B. (2005). The PIN auxin efflux facilitator network controls growth and patterning in Arabidopsis roots. Nature *433*, 39-44.

Bohmdorfer, G., Sethuraman, S., Rowley, M.J., Krzyszton, M., Rothi, M.H., Bouzit, L., and Wierzbicki, A.T. (2016). Long non-coding RNA produced by RNA polymerase V determines boundaries of heterochromatin. Elife *5:e19092*.

Boruc, J., and Van Damme, D. (2015). Endomembrane trafficking overarching cell plate formation. Curr Opin Plant Biol *28*, 92-98.

Boudolf, V., Barroco, R., Engler Jde, A., Verkest, A., Beeckman, T., Naudts, M., Inze, D., and De Veylder, L. (2004). B1-type cyclin-dependent kinases are essential for the formation of stomatal complexes in Arabidopsis thaliana. Plant Cell *16*, 945-955.

Boudolf, V., Lammens, T., Boruc, J., Van Leene, J., Van Den Daele, H., Maes, S., Van Isterdael, G., Russinova, E., Kondorosi, E., Witters, E*., et al.* (2009). CDKB1;1 forms a functional complex with CYCA2;3 to suppress endocycle onset. Plant Physiol *150*, 1482-1493.

Breuer, C., Ishida, T., and Sugimoto, K. (2010). Developmental control of endocycles and cell growth in plants. Curr Opin Plant Biol *13*, 654-660.

Bucher, E., Reinders, J., and Mirouze, M. (2012). Epigenetic control of transposon transcription and mobility in Arabidopsis. Curr Opin Plant Biol *15*, 503-510.

Cadoret, J.C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A *105*, 15837-15842.

Cardenas, P.D., Gajardo, H.A., Huebert, T., Parkin, I.A., Iniguez-Luy, F.L., and Federico, M.L. (2012). Retention of triplicated phytoene synthase (PSY) genes in Brassica napus L. and its diploid progenitors during the evolution of the Brassiceae. Theor Appl Genet *124*, 1215-1228.

Caro, E., Castellano, M.M., and Gutierrez, C. (2007). A chromatin link that couples cell division to root epidermis patterning in Arabidopsis. Nature *447*, 213-217.

Caro, E., and Gutierrez, C. (2007). A green GEM: intriguing analogies with animal geminin. Trends Cell Biol *17*, 580-585.

Casamitjana-Martinez, E., Hofhuis, H.F., Xu, J., Liu, C.M., Heidstra, R., and Scheres, B. (2003). Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of Arabidopsis root meristem maintenance. Curr Biol *13*, 1435-1441.

Castellano, M.M., Boniotti, M.B., Caro, E., Schnittger, A., and Gutierrez, C. (2004). DNA replication licensing affects cell proliferation or endoreplication in a cell type-specific manner. Plant Cell *16*, 2380-2393.

Castellano, M.M., del Pozo, J.C., Ramirez-Parra, E., Brown, S., and Gutierrez, C. (2001). Expression and stability of Arabidopsis CDC6 are associated with endoreplication. Plant Cell *13*, 2671-2686.

Cavrak, V.V., Lettner, N., Jamge, S., Kosarewicz, A., Bayer, L.M., and Mittelsten Scheid, O. (2014). How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet *10*, e1004115.

Cayrou, C., Ballester, B., Peiffer, I., Fenouil, R., Coulombe, P., Andrau, J.C., van Helden, J., and Mechali, M. (2015). The chromatin environment shapes DNA replication origin organization and defines origin classes. Genome Res *25*, 1873-1885.

Cayrou, C., Coulombe, P., Puy, A., Rialle, S., Kaplan, N., Segal, E., and Mechali, M. (2012). New insights into replication origin characteristics in metazoans. Cell Cycle *11*, 658-667.

Cayrou, C., Coulombe, P., Vigneron, A., Stanojcic, S., Ganier, O., Peiffer, I., Rivals, E., Puy, A., Laurent-Chabalier, S., Desprat, R*., et al.* (2011). Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res *21*, 1438-1449.

Chakraborty, A., Shen, Z., and Prasanth, S.G. (2011). "ORCanization" on heterochromatin: linking DNA replication initiation to chromatin organization. Epigenetics *6*, 665-670.

Chang, F., May, C.D., Hoggard, T., Miller, J., Fox, C.A., and Weinreich, M. (2011). High-resolution analysis of four efficient yeast replication origins reveals new insights into the ORC and putative MCM binding elements. Nucleic Acids Res *39*, 6523-6535.

Chenais, B., Caruso, A., Hiard, S., and Casse, N. (2012). The impact of transposable elements on eukaryotic genomes: from genome size increase to genetic adaptation to stressful environments. Gene *509*, 7-15.

Cheng, C.Y., Krishnakumar, V., Chan, A.P., Thibaud-Nissen, F., Schobel, S., and Town, C.D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. Plant J *89*, 789-804.

Chuang, R.Y., and Kelly, T.J. (1999). The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks. Proc Natl Acad Sci U S A *96*, 2656-2661.

Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat Rev Genet *18*, 71-86.

Chupeau, M.C., Granier, F., Pichon, O., Renou, J.P., Gaudin, V., and Chupeau, Y. (2013). Characterization of the early events leading to totipotency in an Arabidopsis protoplast liquid culture by temporal transcript profiling. Plant Cell *25*, 2444-2463.

Churchman, M.L., Brown, M.L., Kato, N., Kirik, V., Hulskamp, M., Inze, D., De Veylder, L., Walker, J.D., Zheng, Z., Oppenheimer, D.G.*, et al.* (2006). SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana. Plant Cell *18*, 3145-3157.

Colon-Carmona, A., You, R., Haimovitch-Gal, T., and Doerner, P. (1999). Technical advance: spatio-temporal analysis of mitotic activity with a labile cyclin-GUS fusion protein. Plant J *20*, 503-508.

Comoglio, F., Schlumpf, T., Schmid, V., Rohs, R., Beisel, C., and Paro, R. (2015). High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins. Cell Rep *11*, 821-834.

Costas, C., de la Paz Sanchez, M., Stroud, H., Yu, Y., Oliveros, J.C., Feng, S., Benguria, A., Lopez-Vidriero, I., Zhang, X., Solano, R.*, et al.* (2011). Genome-wide mapping of Arabidopsis thaliana origins of DNA replication and their associated epigenetic marks. Nature Struc Mol Biol *18*, 395-400.

Coulombe, P., Gregoire, D., Tsanov, N., and Mechali, M. (2013). A spontaneous Cdt1 mutation in 129 mouse strains reveals a regulatory domain restraining replication licensing. Nat Commun *4*, 2065.

Cruz-Ramirez, A., Diaz-Trivino, S., Wachsman, G., Du, Y., Arteaga-Vazquez, M., Zhang, H., Benjamins, R., Blilou, I., Neef, A.B., Chandler, V.*, et al.* (2013). A SCARECROW-RETINOBLASTOMA protein network controls protective quiescence in the Arabidopsis root stem cell organizer. PLoS Biol *11*, e1001724.

Cui, H., and Benfey, P.N. (2009). Interplay between SCARECROW, GA and LIKE HETEROCHROMATIN PROTEIN 1 in ground tissue patterning in the Arabidopsis root. Plant J *58*, 1016-1027.

de la Paz Sanchez, M., and Gutierrez, C. (2009). Arabidopsis ORC1 is a PHD-containing H3K4me3 effector that regulates transcription. Proc Natl Acad Sci U S A *106*, 2065-2070.

de Munnik, S.A., Bicknell, L.S., Aftimos, S., Al-Aama, J.Y., van Bever, Y., Bober, M.B., Clayton-Smith, J., Edrees, A.Y., Feingold, M., Fryer, A.*, et al.* (2012). Meier-Gorlin syndrome genotype-phenotype studies: 35 individuals with pre-replication complex gene mutations and 10 without molecular diagnosis. Eur J Hum Genet *20*, 598-606.

De Rybel, B., Mahonen, A.P., Helariutta, Y., and Weijers, D. (2016). Plant vascular development: from early specification to differentiation. Nat Rev Mol Cell Biol *17*, 30-40.

De Schutter, K., Joubes, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S., Inze, D.*, et al.* (2007). Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. Plant Cell *19*, 211-225.

De Veylder, L., Beeckman, T., Beemster, G.T., Krols, L., Terras, F., Landrieu, I., van der Schueren, E., Maes, S., Naudts, M., and Inze, D. (2001). Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis. Plant Cell *13*, 1653-1668.

Deegan, T.D., Yeeles, J.T., and Diffley, J.F. (2016). Phosphopeptide binding by Sld3 links Dbf4-dependent kinase to MCM replicative helicase activation. EMBO J *35*, 961-973.

del Pozo, J.C., Boniotti, M.B., and Gutierrez, C. (2002). Arabidopsis E2Fc functions in cell division and is degraded by the ubiquitin-SCF(AtSKP2) pathway in response to light. Plant Cell *14*, 3057-3071.

del Pozo, J.C., Diaz-Trivino, S., Cisneros, N., and Gutierrez, C. (2006). The balance between cell division and endoreplication depends on E2FC-DPB, transcription factors regulated by the ubiquitin-SCFSKP2A pathway in Arabidopsis. Plant Cell *18*, 2224-2235.

Dello Ioio, R., Nakamura, K., Moubayidin, L., Perilli, S., Taniguchi, M., Morita, M.T., Aoyama, T., Costantino, P., and Sabatini, S. (2008). A genetic framework for the control of cell division and differentiation in the root meristem. Science *322*, 1380-1384.

Demidov, D., Van Damme, D., Geelen, D., Blattner, F.R., and Houben, A. (2005). Identification and dynamics of two classes of aurora-like kinases in Arabidopsis and other plants. Plant Cell *17*, 836-848.

DePamphilis, M.L. (2005). Cell cycle dependent regulation of the origin recognition complex. Cell Cycle *4*, 70-79.

Deragon, J.M., Casacuberta, J.M., and Panaud, O. (2008). Plant transposable elements. Genome Dyn *4*, 69-82.

Desvoyes, B., Fernandez-Marcos, M., Sequeira-Mendes, J., Otero, S., Vergara, Z., and Gutierrez, C. (2014). Looking at plant cell cycle from the chromatin window. Front Plant Sci *5*, 369.

Desvoyes, B., Vergara, Z., Sequeira-Mendes, J., Madeira, S., Gutierrez, C. (2017). A rapid and efficient ChIP protocol to profile chromatin binding proteins and epigenetic modifications in bulk Arabidopsis tissue. Methods Mol Biol *In press*.

Dharmasiri, N., Dharmasiri, S., Jones, A.M., and Estelle, M. (2003). Auxin action in a cell-free system. Curr Biol *13*, 1418-1422.

Diaz-Trivino, S. (2005). Identificación y caracterización del complejo ORC y estudio de la iniciación de la replicación en una región del cromosoma II de arabidopsis thaliana (Madrid: UAM).

Diaz-Trivino, S., Castellano, M.M., Sanchez, M.P., Ramirez-Parra, E., Desvoyes, B., and Gutierrez, C. (2005). The genes encoding Arabidopsis ORC subunits are E2F targets and the two ORC1 genes are differently expressed in proliferating and endoreplicating cells. Nucleic Acids Res *33*, 5404-5414.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376-380.

Domenichini, S., Benhamed, M., De Jaeger, G., Van De Slijke, E., Blanchet, S., Bourge, M., De Veylder, L., Bergounioux, C., and Raynaud, C. (2012). Evidence for a role of Arabidopsis CDT1 proteins in gametophyte development and maintenance of genome integrity. Plant Cell *24*, 2779-2791.

Donti, T.R., Datta, S., Sandoval, P.Y., and Kapler, G.M. (2009). Differential targeting of Tetrahymena ORC to ribosomal DNA and non-rDNA replication origins. EMBO J *28*, 223-233.

Drouaud, J., Camilleri, C., Bourguignon, P.Y., Canaguier, A., Berard, A., Vezon, D., Giancola, S., Brunel, D., Colot, V., Prum, B*., et al.* (2006). Variation in crossing-over rates across chromosome 4 of Arabidopsis thaliana reveals the presence of meiotic recombination "hot spots". Genome Res *16*, 106-114.

Drouaud, J., Mercier, R., Chelysheva, L., Berard, A., Falque, M., Martin, O., Zanni, V., Brunel, D., and Mezard, C. (2007). Sex-specific crossover distributions and variations in interference level along Arabidopsis thaliana chromosome 4. PLoS Genet *3*, e106.

Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat Rev Mol Cell Biol *16*, 519-532.

Dueber, E.L., Corn, J.E., Bell, S.D., and Berger, J.M. (2007). Replication origin recognition and deformation by a heterodimeric archaeal Orc1 complex. Science *317*, 1210-1213.

Duncker, B.P., Chesnokov, I.N., and McConkey, B.J. (2009). The origin recognition complex protein family. Genome Biol *10*, 214.

Edgar, B.A., Zielke, N., and Gutierrez, C. (2014). Endocycles: a recurrent evolutionary innovation for post-mitotic cell growth. Nat Rev Mol Cell Biol *15*, 197-210.

Evrin, C., Fernandez-Cid, A., Riera, A., Zech, J., Clarke, P., Herrera, M.C., Tognetti, S., Lurz, R., and Speck, C. (2014). The ORC/Cdc6/MCM2-7 complex facilitates MCM2-7 dimerization during prereplicative complex formation. Nucleic Acids Res *42*, 2257-2269.

Exner, V., Taranto, P., Schonrock, N., Gruissem, W., and Hennig, L. (2006). Chromatin assembly factor CAF-1 is required for cellular differentiation during plant development. Development *133*, 4163-4172.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol Cell *55*, 694-707.

Feng, W., and Michaels, S.D. (2015). Accessing the Inaccessible: The Organization, Transcription, Replication, and Repair of Heterochromatin in Plants. Annu Rev Genet *49*, 439-459.

Fernandez-Marcos, M., Desvoyes, B., Manzano, C., Liberman, L.M., Benfey, P.N., Del Pozo, J.C., and Gutierrez, C. (2017). Control of Arabidopsis lateral root primordium boundaries by MYB36. New Phytol *213*, 105-112.

Fragkos, M., Ganier, O., Coulombe, P., and Mechali, M. (2015). DNA replication origin activation in space and time. Nat Rev Mol Cell Biol *16*, 360-374.

Francis, L.I., Randell, J.C., Takara, T.J., Uchima, L., and Bell, S.P. (2009). Incorporation into the prereplicative complex activates the Mcm2-7 helicase for Cdc7-Dbf4 phosphorylation. Genes Dev *23*, 643-654.

Fransz, P., ten Hoopen, R., and Tessadori, F. (2006). Composition and formation of heterochromatin in Arabidopsis thaliana. Chromosome Res *14*, 71-82.

Frigola, J., Remus, D., Mehanna, A., and Diffley, J.F. (2013). ATPase-dependent quality control of DNA replication origin licensing. Nature *495*, 339-343.

Fülöp, K., Tarayre, S., Kelemen, Z., Horvath, G., Kevei, Z., Nikovics, K., Bako, L., Brown, S., Kondorosi, A., and Kondorosi, E. (2005). Arabidopsis anaphase-promoting complexes: multiple activators and wide range of substrates might keep APC perpetually busy. Cell Cycle *4*, 1084-1092.

Fultz, D., Choudury, S.G., and Slotkin, R.K. (2015). Silencing of active transposable elements in plants. Curr Opin Plant Biol *27*, 67-76.

Gagne, J.M., Downes, B.P., Shiu, S.H., Durski, A.M., and Vierstra, R.D. (2002). The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in Arabidopsis. Proc Natl Acad Sci U S A *99*, 11519-11524.

Galbraith, D.W., Harkins, K.R., and Knapp, S. (1991). Systemic endopolyploidy in *Arabopsis thaliana*. Plant Physiol *96*, 985-989.

Galichet, A., and Gruissem, W. (2006). Developmentally controlled farnesylation modulates AtNAP1;1 function in cell proliferation and cell expansion during Arabidopsis leaf development. Plant Physiol *142*, 1412-1426.

Gaudier, M., Schuwirth, B.S., Westcott, S.L., and Wigley, D.B. (2007). Structural basis of DNA replication origin recognition by an ORC protein. Science *317*, 1213-1216.

Gerbi, S.A., and Bielinsky, A.K. (1997). Replication initiation point mapping. Methods *13*, 271-280.

Godoy, P.D., Nogueira-Junior, L.A., Paes, L.S., Cornejo, A., Martins, R.M., Silber, A.M., Schenkman, S., and Elias, M.C. (2009). Trypanosome prereplication machinery contains a single functional orc1/cdc6 protein, which is typical of archaea. Eukaryot Cell *8*, 1592-1603.

Gomez, M., and Antequera, F. (1999). Organization of DNA replication origins in the fission yeast genome. EMBO J *18*, 5683-5690.

Gomez, M., and Antequera, F. (2008). Overreplication of short DNA regions during S phase in human cells. Genes Dev *22*, 375-385.

Gomez, M., and Brockdorff, N. (2004). Heterochromatin on the inactive X chromosome delays replication timing without affecting origin usage. Proc Natl Acad Sci U S A *101*, 6923-6928.

Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: Spatial Control Over Gene Expression. Trends Genet *32*, 485-495.

Graham, T., and Boissinot, S. (2006). The genomic distribution of L1 elements: the role of insertion bias and natural selection. J Biomed Biotechnol *2006*, 75327.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. Mol Cell *55*, 678-693.

Gupta, A., Mehra, P., Deshmukh, A., Dar, A., Mitra, P., Roy, N., and Dhar, S.K. (2009). Functional dissection of the catalytic carboxyl-terminal domain of origin recognition complex subunit 1 (PfORC1) of the human malaria parasite Plasmodium falciparum. Eukaryot Cell *8*, 1341-1351.

Gupta, A., Mehra, P., and Dhar, S.K. (2008). Plasmodium falciparum origin recognition complex subunit 5: functional characterization and role in DNA replication foci formation. Mol Microbiol *69*, 646-665.

Gusti, A., Baumberger, N., Nowack, M., Pusch, S., Eisler, H., Potuschak, T., De Veylder, L., Schnittger, A., and Genschik, P. (2009). The Arabidopsis thaliana F-box protein FBL17 is essential for progression through the second mitosis during pollen development. PLoS One *4*, e4780.

Gutierrez, C. (2009). The Arabidopsis cell division cycle. In Arabidopsis Book (ASPB), p. 7:e0120.

Gutierrez, C. (2016). 25 Years of Cell Cycle Research: What's Ahead? Trends Plant Sci *21*, 823-833.

Gutierrez, C., Desvoyes, B., Vergara, Z., Otero, S., and Sequeira-Mendes, J. (2016). Links of genome replication, transcriptional silencing and chromatin dynamics. Curr Opin Plant Biol *34*, 92-99.

Gutierrez, C., Ramirez-Parra, E., Castellano, M.M., and del Pozo, J.C. (2002). G(1) to S transition: more than a cell cycle engine switch. Curr Opin Plant Biol *5*, 480-486.

Haecker, A., Gross-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann, M., and Laux, T. (2004). Expression dynamics of WOX genes mark cell fate decisions during early embryonic patterning in Arabidopsis thaliana. Development *131*, 657-668.

Hakenjos, J.P., Richter, R., Dohmann, E.M., Katsiarimpa, A., Isono, E., and Schwechheimer, C. (2011). MLN4924 is an efficient inhibitor of NEDD8 conjugation in plants. Plant Physiol *156*, 527-536.

Hale, C.J., Potok, M.E., Lopez, J.A., Do, T., Liu, A., Gallego-Bartolome, J., Michaels, S.D., and Jacobsen, S.E. (2016). Identification of multiple proteins coupliing transcriptional gene-silencing to genome stability in Arabidopsis thaliana. PLoS Genet.

Hayashi, K., Hasegawa, J., and Matsunaga, S. (2013). The boundary of the meristematic and elongation zones in roots: endoreduplication precedes rapid cell expansion. Sci Rep *3*, 2723.

Heidstra, R., and Sabatini, S. (2014). Plant and animal stem cells: similar yet different. Nat Rev Mol Cell Biol *15*, 301-312.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Helariutta, Y., Fukaki, H., Wysocka-Diller, J., Nakajima, K., Jung, J., Sena, G., Hauser, M.T., and Benfey, P.N. (2000). The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. Cell *101*, 555-567.

Henaff, E., Vives, C., Desvoyes, B., Chaurasia, A., Payet, J., Gutierrez, C., and Casacuberta, J.M. (2014). Extensive amplification of the E2F transcription factor binding sites by transposons during evolution of Brassica species. Plant J *77*, 852-862.

Herr, A.J., Jensen, M.B., Dalmay, T., and Baulcombe, D.C. (2005). RNA polymerase IV directs silencing of endogenous DNA. Science *308*, 118-120.

Herridge, R.P., Day, R.C., and Macknight, R.C. (2014). The role of the MCM2-7 helicase complex during Arabidopsis seed development. Plant Mol Biol *86*, 69-84.

Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S., and Johnson, F.B. (2008). Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. Nucleic Acids Res *36*, 144-156.

Heyman, J., Cools, T., Vandenbussche, F., Heyndrickx, K.S., Van Leene, J., Vercauteren, I., Vanderauwera, S., Vandepoele, K., De Jaeger, G., Van Der Straeten, D*., et al.* (2013). ERF115 controls root quiescent center cell division and stem cell replenishment. Science *342*, 860-863.

Hickman, M.A., and Rusche, L.N. (2010). Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication. Proc Natl Acad Sci U S A *107*, 19384-19389.

Hirsch, C.D., and Springer, N.M. (2017). Transposable element influences on gene expression in plants. Biochim Biophys Acta *1860*, 157-165.

Hirsch, S., and Oldroyd, G.E. (2009). GRAS-domain transcription factors that regulate plant development. Plant Signal Behav *4*, 698-700.

Hoshina, S., Yura, K., Teranishi, H., Kiyasu, N., Tominaga, A., Kadoma, H., Nakatsuka, A., Kunichika, T., Obuse, C., and Waga, S. (2013). Human origin recognition complex binds preferentially to G-quadruplex-preferable RNA and single-stranded DNA. J Biol Chem *288*, 30161-30171.

Hou, Z., Bernstein, D.A., Fox, C.A., and Keck, J.L. (2005). Structural basis of the Sir1-origin recognition complex interaction in transcriptional silencing. Proc Natl Acad Sci U S A *102*, 8489-8494.

Hower, V., Evans, S.N., and Pachter, L. (2011). Shape-based peak identification for ChIP-Seq. BMC Bioinformatics *12*, 15.

Hudson, M.E., and Quail, P.H. (2003). Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. Plant Physiol *133*, 1605-1616.

Hyrien, O., and Mechali, M. (1993). Chromosomal replication initiates and terminates at random sequences but at regular intervals in the ribosomal DNA of Xenopus early embryos. EMBO J *12*, 4511-4520.

Iizuka, M., and Stillman, B. (1999). Histone acetyltransferase HBO1 interacts with the ORC1 subunit of the human initiator protein. J Biol Chem *274*, 23027-23034.

Inze, D., and De Veylder, L. (2006). Cell cycle regulation in plant development. Annu Rev Genet *40*, 77-105.

Ivanov, V.B., and Dubrovsky, J.G. (2013). Longitudinal zonation pattern in plant roots: conflicts and solutions. Trends Plant Sci *18*, 237-243.

Jackson, J.P., Lindroth, A.M., Cao, X., and Jacobsen, S.E. (2002). Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. Nature *416*, 556-560.

Jacob, Y., Feng, S., LeBlanc, C.A., Bernatavichute, Y.V., Stroud, H., Cokus, S., Johnson, L.M., Pellegrini, M., Jacobsen, S.E., and Michaels, S.D. (2009). ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol *16*, 763-768.

Jacob, Y., Stroud, H., Leblanc, C., Feng, S., Zhuo, L., Caro, E., Hassel, C., Gutierrez, C., Michaels, S.D., and Jacobsen, S.E. (2010). Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature *466*, 987-991.

Jefferson, R.A., Kavanagh, T.A., and Bevan, M.W. (1987). GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. Embo J *6*, 3901-3907.

Kara, N., Hossain, M., Prasanth, S.G., and Stillman, B. (2015). Orc1 Binding to Mitotic Chromosomes Precedes Spatial Patterning during G1 Phase and Assembly of the Origin Recognition Complex in Human Cells. J Biol Chem *290*, 12355-12369.

Kawabe, A., Matsunaga, S., Nakagawa, K., Kurihara, D., Yoneda, A., Hasezawa, S., Uchiyama, S., and Fukui, K. (2005). Characterization of plant Aurora kinases during mitosis. Plant Mol Biol *58*, 1-13.

Kazazian, H.H., Jr. (2000). Genetics. L1 retrotransposons shape the mammalian genome. Science *289*, 1152-1153.

Kejnovsky, E., Tokan, V., and Lexa, M. (2015). Transposable elements and G-quadruplexes. Chromosome Res *23*, 615-623.

Kopytova, D., Popova, V., Kurshakova, M., Shidlovskii, Y., Nabirochkina, E., Brechalov, A., Georgiev, G., and Georgieva, S. (2016). ORC interacts with THSC/TREX-2 and its subunits promote Nxf1 association with mRNP and mRNA export in Drosophila. Nucleic Acids Res *44*, 4920-4933.

Korbie, D.J., and Mattick, J.S. (2008). Touchdown PCR for increased specificity and sensitivity in PCR amplification. Nat Protoc *3*, 1452-1456.

Koren, A., and McCarroll, S.A. (2014). Random replication of the inactive X chromosome. Genome Res *24*, 64-69.

Kumar, D., Mukherji, A., and Saha, S. (2008). Expression and subcellular localization of ORC1 in Leishmania major. Biochem Biophys Res Commun *375*, 74-79.

Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. Nature *484*, 115–119.

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M.*, et al.* (2012). The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res *40*, D1202-1210.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat Rev Genet *11*, 204-220.

Lee, T.J., Pascuzzi, P.E., Settlage, S.B., Shultz, R.W., Tanurdzic, M., Rabinowicz, P.D., Menges, M., Zheng, P., Main, D., Murray, J.A.*, et al.* (2010). Arabidopsis thaliana chromosome 4 replicates in two phases that correlate with chromatin state. PLoS Genet *6*, e1000982.

Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J., and Schubert, I. (2006). Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. Plant Cell *18*, 2443-2451.

Li, L., and Clevers, H. (2010). Coexistence of quiescent and active adult stem cells in mammals. Science *327*, 542-545.

Li, S., Yang, Z., Du, X., Liu, R., Wilkinson, A.W., Gozani, O., Jacobsen, S.E., Patel, D.J., and Du, J. (2016). Structural Basis for the Unique Multivalent Readout of Unmodified H3 Tail by Arabidopsis ORC1b BAH-PHD Cassette. Structure *24*, 486-494.

Liberman, L.M., Sparks, E.E., Moreno-Risueno, M.A., Petricka, J.J., and Benfey, P.N. (2015). MYB36 regulates the transition from proliferation to differentiation in the Arabidopsis root. Proc Natl Acad Sci U S A.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D*., et al.* (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471-476.

Lisch, D. (2013). How important are transposons for plant evolution? Nat Rev Genet *14*, 49-61.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Liu, C., Wang, C., Wang, G., Becker, C., Zaidem, M., and Weigel, D. (2016). Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. Genome Res *26*, 1057-1068.

Liu, C., and Weigel, D. (2015). Chromatin in 3D: progress and prospects for plants. Genome Biol *16*, 170.

Lombrana, R., Alvarez, A., Fernandez-Justel, J.M., Almeida, R., Poza-Carrion, C., Gomes, F., Calzada, A., Requena, J.M., and Gomez, M. (2016). Transcriptionally Driven DNA Replication Program of the Human Parasite Leishmania major. Cell Rep *16*, 1774-1786.

Maiorano, D., Moreau, J., and Mechali, M. (2000). XCDT1 is required for the assembly of pre-replicative complexes in Xenopus laevis. Nature *404*, 622-625.

Manova, V., and Gruszka, D. (2015). DNA damage and repair in plants - from models to crops. Front Plant Sci *6*, 885.

Marahrens, Y., and Stillman, B. (1992). A yeast chromosomal origin of DNA replication defined by multiple functional elements. Science *255*, 817-823.

Masai, H., Matsumoto, S., You, Z., Yoshizawa-Sugata, N., and Oda, M. (2010). Eukaryotic chromosome DNA replication: where, when, and how? Annu Rev Biochem *79*, 89-130.

Matzke, M.A., and Mosher, R.A. (2014). RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. Nat Rev Genet *15*, 394-408.

McClintock, B. (1944). The Relation of Homozygous Deficiencies to Mutations and Allelic Series in Maize. Genetics *29*, 478-502.

McCue, A.D., Panda, K., Nuthikattu, S., Choudury, S.G., Thomas, E.N., and Slotkin, R.K. (2015). ARGONAUTE 6 bridges transposable element mRNA-derived siRNAs to the establishment of DNA methylation. Embo J *34*, 20-35.

Mechali, M., Yoshida, K., Coulombe, P., and Pasero, P. (2013). Genetic and epigenetic determinants of DNA replication origins, position and activation. Curr Opin Genet Dev *23*, 124-131.

Mendez, J., Zou-Yang, X.H., Kim, S.Y., Hidaka, M., Tansey, W.P., and Stillman, B. (2002). Human origin recognition complex large subunit is degraded by ubiquitin-mediated proteolysis after initiation of DNA replication. Mol Cell *9*, 481-491.

Menges, M., and Murray, J.A. (2002). Synchronous Arabidopsis suspension cultures for analysis of cell-cycle gene activity. Plant J *30*, 203-212.

Mita, P., and Boeke, J.D. (2016). How retrotransposons shape genome regulation. Curr Opin Genet Dev *37*, 90-100.

Miyake, Y., Mizuno, T., Yanagi, K., and Hanaoka, F. (2005). Novel splicing variant of mouse Orc1 is deficient in nuclear translocation and resistant for proteasome-mediated degradation. J Biol Chem *280*, 12643-12652.

Mohammad, M.M., Donti, T.R., Sebastian Yakisich, J., Smith, A.G., and Kapler, G.M. (2007). Tetrahymena ORC contains a ribosomal RNA fragment that participates in rDNA origin recognition. EMBO J *26*, 5048-5060.

Morao, A.K., Bouyer, D., and Roudier, F. (2016). Emerging concepts in chromatin-level regulation of plant cell differentiation: timing, counting, sensing and maintaining. Curr Opin Plant Biol *34*, 27-34.

Mozo, T., Fischer, S., Meier-Ewert, S., Lehrach, H., and Altmann, T. (1998). Use of the IGF BAC library for physical mapping of the Arabidopsis thaliana genome. Plant J *16*, 377-384.

Muller, S., and Jurgens, G. (2016). Plant cytokinesis-No ring, no constriction but centrifugal construction of the partitioning membrane. Semin Cell Dev Biol *53*, 10-18.

Munoz, S., and Mendez, J. (2017). DNA replication stress: from molecular mechanisms to human disease. Chromosoma *126*, 1-15.

Muramatsu, S., Hirai, K., Tak, Y.S., Kamimura, Y., and Araki, H. (2010). CDK-dependent complex formation between replication proteins Dpb11, Sld2, Pol (epsilon}, and GINS in budding yeast. Genes Dev *24*, 602-612.

Nakagawa, T., Suzuki, T., Murata, S., Nakamura, S., Hino, T., Maeo, K., Tabata, R., Kawai, T., Tanaka, K., Niwa, Y*., et al.* (2007). Improved Gateway binary vectors: high-performance vectors for creation of fusion constructs in transgenic analysis of plants. Biosci Biotechnol Biochem *71*, 2095-2100.

Nakajima, K., Sena, G., Nawy, T., and Benfey, P.N. (2001). Intercellular movement of the putative transcription factor SHR in root patterning. Nature *413*, 307-311.

Naumann, K., Fischer, A., Hofmann, I., Krauss, V., Phalke, S., Irmler, K., Hause, G., Aurich, A.C., Dorn, R., Jenuwein, T*., et al.* (2005). Pivotal role of AtSUVH2 in heterochromatic histone methylation and gene silencing in Arabidopsis. EMBO J *24*, 1418-1429.

Negi, P., Rai, A.N., and Suprasanna, P. (2016). Moving through the Stressed Genome: Emerging Regulatory Roles for Transposons in Plant Stress Response. Front Plant Sci *7*, 1448.

Neuwald, A.F., Aravind, L., Spouge, J.L., and Koonin, E.V. (1999). AAA+: A class of chaperone-like ATPases associated with the assembly, operation, and disassembly of protein complexes. Genome Res *9*, 27-43.

Newlon, C.S., and Theis, J.F. (1993). The structure and function of yeast ARS elements. Curr Opin Genet Dev *3*, 752-758.

Ngo, Q.A., Baroux, C., Guthorl, D., Mozerov, P., Collinge, M.A., Sundaresan, V., and Grossniklaus, U. (2012). The Armadillo repeat gene ZAK IXIK promotes Arabidopsis early embryo and endosperm development through a distinctive gametophytic maternal effect. Plant Cell *24*, 4026-4043.

Nguyen, V.Q., Co, C., and Li, J.J. (2001). Cyclin-dependent kinases prevent DNA re-replication through multiple mechanisms. Nature *411*, 1068-1073.

Nikolov, I., and Taddei, A. (2016). Linking replication stress with heterochromatin formation. Chromosoma *125*, 523-533.

Noir, S., Marrocco, K., Masoud, K., Thomann, A., Gusti, A., Bitrian, M., Schnittger, A., and Genschik, P. (2015). The Control of Arabidopsis thaliana Growth by Cell Proliferation and Endoreplication Requires the F-Box Protein FBL17. Plant Cell *27*, 1461-1476.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J*., et al.* (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381-385.

Oda, H., Okamoto, I., Murphy, N., Chu, J., Price, S.M., Shen, M.M., Torres-Padilla, M.E., Heard, E., and Reinberg, D. (2009). Monomethylation of histone H4-lysine 20 is involved in chromosome structure and stability and is essential for mouse development. Mol Cell Biol *29*, 2278-2295.

Onodera, Y., Haag, J.R., Ream, T., Costa Nunes, P., Pontes, O., and Pikaard, C.S. (2005). Plant nuclear RNA polymerase IV mediates siRNA and DNA methylation-dependent heterochromatin formation. Cell *120*, 613-622.

Otero, S., Desvoyes, B., Peiro, R., and Gutierrez, C. (2016). Histone H3 dynamics uncovers domains with distinct proliferation potential in the Arabidopsis root. Plant Cell *28*, 1361-1371.

Pak, D.T., Pflumm, M., Chesnokov, I., Huang, D.W., Kellum, R., Marr, J., Romanowski, P., and Botchan, M.R. (1997). Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. Cell *91*, 311-323.

Parfrey, L.W., Lahr, D.J., Knoll, A.H., and Katz, L.A. (2011). Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A *108*, 13624-13629.

Park, S.Y., and Asano, M. (2008). The origin recognition complex is dispensable for endoreplication in Drosophila. Proc Natl Acad Sci U S A *105*, 12343-12348.

Pecinka, A., Schubert, V., Meister, A., Kreth, G., Klatte, M., Lysak, M.A., Fuchs, J., and Schubert, I. (2004). Chromosome territory arrangement and homologous pairing in nuclei of Arabidopsis thaliana are predominantly random except for NOR-bearing chromosomes. Chromosoma *113*, 258-269.

Petersen, B.O., Lukas, J., Sorensen, C.S., Bartek, J., and Helin, K. (1999). Phosphorylation of mammalian CDC6 by cyclin A/CDK2 regulates its subcellular localization. EMBO J *18*, 396-410.

Petersen, B.O., Wagener, C., Marinoni, F., Kramer, E.R., Melixetian, M., Lazzerini Denchi, E., Gieffers, C., Matteucci, C., Peters, J.M., and Helin, K. (2000). Cell cycle- and cell growth-regulated proteolysis of mammalian CDC6 is dependent on APC-CDH1. Genes Dev *14*, 2330-2343.

Petrasek, J., and Friml, J. (2009). Auxin transport routes in plant development. Development *136*, 2675-2688.

Petricka, J.J., Winter, C.M., and Benfey, P.N. (2012). Control of Arabidopsis Root Development. Annu Rev Plant Biol *63*, 563–590.

Picard, F., Cadoret, J.C., Audit, B., Arneodo, A., Alberti, A., Battail, C., Duret, L., and Prioleau, M.N. (2014). The spatiotemporal program of DNA replication is associated with specific combinations of chromatin marks in human cells. PLoS Genet *10*, e1004282.

Pikaard, C.S., and Mittelsten Scheid, O. (2014). Epigenetic regulation in plants. Cold Spring Harb Perspect Biol *6*, a019315.

Pontier, D., Yahubyan, G., Vega, D., Bulski, A., Saez-Vasquez, J., Hakimi, M.A., Lerbs-Mache, S., Colot, V., and Lagrange, T. (2005). Reinforcement of silencing at transposons and highly repeated sequences requires the concerted action of two distinct RNA polymerases IV in Arabidopsis. Genes Dev *19*, 2030-2040.

Prasanth, S.G., Shen, Z., Prasanth, K.V., and Stillman, B. (2010). Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization. Proc Natl Acad Sci U S A *107*, 15093-15098.

Prunet, N., and Meyerowitz, E.M. (2016). Genetics and plant development. C R Biol *339*, 240-246.

Puchta, H. (2017). Applying CRISPR/Cas for genome engineering in plants: the best is yet to come. Curr Opin Plant Biol *36*, 1-8.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Curr Protoc Bioinformatics *47*, 11.12.11-34.

Ramirez-Parra, E., and Gutierrez, C. (2007). The many faces of chromatin assembly factor 1. Trends Plant Sci *12*, 570-576.

Remus, D., Beuron, F., Tolun, G., Griffith, J.D., Morris, E.P., and Diffley, J.F. (2009). Concerted loading of Mcm2-7 double hexamers around DNA during DNA replication origin licensing. Cell *139*, 719-730.

Remus, D., Blanchette, M., Rio, D.C., and Botchan, M.R. (2005). CDK phosphorylation inhibits the DNA-binding and ATP-hydrolysis activities of the Drosophila origin recognition complex. J Biol Chem *280*, 39740-39751.

Ren, H., Santner, A., del Pozo, J.C., Murray, J.A., and Estelle, M. (2008). Degradation of the cyclin-dependent kinase inhibitor KRP1 is regulated by two different ubiquitin E3 ligases. Plant J *53*, 705-716.

Robinson, N.P., and Bell, S.D. (2005). Origins of DNA replication in the three domains of life. FEBS J *272*, 3757-3766.

Rodriguez-Martinez, M., Pinzon, N., Ghommidh, C., Beyne, E., Seitz, H., Cayrou, C., and Mechali, M. (2017). The gastrula transition reorganizes replication-origin selection in Caenorhabditis elegans. Nat Struct Mol Biol *24*, 290-299.

Sabatini, S., Heidstra, R., Wildwater, M., and Scheres, B. (2003). SCARECROW is involved in positioning the stem cell niche in the Arabidopsis root meristem. Genes Dev *17*, 354-358.

Sarkar, A.K., Luijten, M., Miyashima, S., Lenhard, M., Hashimoto, T., Nakajima, K., Scheres, B., Heidstra, R., and Laux, T. (2007). Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers. Nature *446*, 811-814.

Sasaki, T., Sawado, T., Yamaguchi, M., and Shinomiya, T. (1999). Specification of regions of DNA replication initiation during embryogenesis in the 65-kilobase DNApolalpha-dE2F locus of Drosophila melanogaster. Mol Cell Biol *19*, 547-555.

Schaller, G.E., Bishopp, A., and Kieber, J.J. (2015). The yin-yang of hormones: cytokinin and auxin interactions in plant development. Plant Cell *27*, 44-63.

Scheres, B. (2007). Stem-cell niches: nursery rhymes across kingdoms. Nat Rev Mol Cell Biol *8*, 345-354.

Schotta, G., Sengupta, R., Kubicek, S., Malin, S., Kauer, M., Callen, E., Celeste, A., Pagani, M., Opravil, S., De La Rosa-Velazquez, I.A*., et al.* (2008). A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. Genes Dev *22*, 2048-2061.

Schubert, V., Berr, A., and Meister, A. (2012). Interphase chromatin organisation in Arabidopsis nuclei: constraints versus randomness. Chromosoma *121*, 369-387.

Sequeira-Mendes, J., Araguez, I., Peiro, R., Mendez-Giraldez, R., Zhang, X., Jacobsen, S.E., Bastolla, U., and Gutierrez, C. (2014). The Functional Topography of the Arabidopsis Genome Is Organized in a Reduced Number of Linear Motifs of Chromatin States. Plant Cell *26*, 2351-2366

Sequeira-Mendes, J., Diaz-Uriarte, R., Apedaile, A., Huntley, D., Brockdorff, N., and Gomez, M. (2009). Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet *5*, e1000446.

Sequeira-Mendes, J., and Gutierrez, C. (2015). Links between genome replication and chromatin landscapes. Plant J *83*, 38-51.

Sequeira-Mendes, J., and Gutierrez, C. (2016). Genome architecture: from linear organisation of chromatin to the 3D assembly in the nucleus. Chromosoma *125*, 455-469.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.

Shareef, M.M., King, C., Damaj, M., Badagu, R., Huang, D.W., and Kellum, R. (2001). Drosophila heterochromatin protein 1 (HP1)/origin recognition complex (ORC) protein is associated with HP1 and ORC and functions in heterochromatin-induced silencing. Mol Biol Cell *12*, 1671-1685.

Sheu, Y.J., and Stillman, B. (2006). Cdc7-Dbf4 phosphorylates MCM proteins via a docking site-mediated mechanism to promote S phase progression. Mol Cell *24*, 101-113.

Shibata, E., Kiran, M., Shibata, Y., Singh, S., Kiran, S., and Dutta, A. (2016). Two subunits of human ORC are dispensable for DNA replication and proliferation. Elife *5*.

Sonneville, R., Querenet, M., Craig, A., Gartner, A., and Blow, J.J. (2012). The dynamics of replication licensing in live Caenorhabditis elegans embryos. J Cell Biol *196*, 233-246.

Sozzani, R., Cui, H., Moreno-Risueno, M.A., Busch, W., Van Norman, J.M., Vernoux, T., Brady, S.M., Dewitte, W., Murray, J.A., and Benfey, P.N. (2010). Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. Nature *466*, 128-132.

Speck, C., Chen, Z., Li, H., and Stillman, B. (2005). ATPase-dependent cooperative binding of ORC and Cdc6 to origin DNA. Nat Struct Mol Biol *12*, 965-971.

Speck, C., and Stillman, B. (2007). Cdc6 ATPase activity regulates ORC x Cdc6 stability and the selection of specific DNA sequences as origins of DNA replication. J Biol Chem *282*, 11705-11714.

Spyrou, C., Stark, R., Lynch, A.G., and Tavare, S. (2009). BayesPeak: Bayesian analysis of ChIP-seq data. BMC Bioinformatics *10*, 299.

Stanojcic, S., Lemaitre, J.M., Brodolin, K., Danis, E., and Mechali, M. (2008). In Xenopus egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences. Mol Cell Biol *28*, 5265-5274.

Stevens, R., Mariconti, L., Rossignol, P., Perennes, C., Cella, R., and Bergounioux, C. (2002). Two E2F sites in the Arabidopsis MCM3 promoter have different roles in cell cycle activation and meristematic expression. J Biol Chem *277*, 32978-32984.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat Struct Mol Biol *21*, 64-72.

Stroud, H., Greenberg, M.V., Feng, S., Bernatavichute, Y.V., and Jacobsen, S.E. (2013). Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome. Cell *152*, 352-364.

Sugimoto, N., Kitabayashi, I., Osano, S., Tatsumi, Y., Yugawa, T., Narisawa-Saito, M., Matsukage, A., Kiyono, T., and Fujita, M. (2008). Identification of novel human Cdt1-binding proteins by a proteomics approach: proteolytic regulation by APC/CCdh1. Mol Biol Cell *19*, 1007-1021.

Sugimoto, N., Tatsumi, Y., Tsurumi, T., Matsukage, A., Kiyono, T., Nishitani, H., and Fujita, M. (2004). Cdt1 phosphorylation by cyclin A-dependent kinases negatively regulates its function without affecting geminin binding. J Biol Chem *279*, 19691-19697.

Sun, J., Evrin, C., Samel, S.A., Fernandez-Cid, A., Riera, A., Kawakami, H., Stillman, B., Speck, C., and Li, H. (2013). Cryo-EM structure of a helicase loading intermediate containing ORC-Cdc6-Cdt1-MCM2-7 bound to DNA. Nat Struct Mol Biol *20*, 944-951.

Tanaka, S., and Diffley, J.F. (2002). Interdependent nuclear accumulation of budding yeast Cdt1 and Mcm2-7 during G1 phase. Nat Cell Biol *4*, 198-207.

Tanaka, S., Umemori, T., Hirai, K., Muramatsu, S., Kamimura, Y., and Araki, H. (2007). CDK-dependent phosphorylation of Sld2 and Sld3 initiates DNA replication in budding yeast. Nature *445*, 328-332.

Tanurdzic, M., Vaughn, M.W., Jiang, H., Lee, T.J., Slotkin, R.K., Sosinski, B., Thompson, W.F., Doerge, R.W., and Martienssen, R.A. (2008). Epigenomic consequences of immortalized plant cell suspension culture. PLoS Biol *6*, 2880-2895.

Tarayre, S., Vinardell, J.M., Cebolla, A., Kondorosi, A., and Kondorosi, E. (2004). Two classes of the CDh1-type activators of the anaphase-promoting complex in plants: novel functional domains and distinct regulation. Plant Cell *16*, 422-434.

Tariq, M., Saze, H., Probst, A.V., Lichota, J., Habu, Y., and Paszkowski, J. (2003). Erasure of CpG methylation in Arabidopsis alters patterns of histone H3 methylation in heterochromatin. Proc Natl Acad Sci U S A *100*, 8823-8827.

Tian, Q., Uhlir, N.J., and Reed, J.W. (2002). Arabidopsis SHY2/IAA3 inhibits auxin-regulated gene expression. Plant Cell *14*, 301-319.

Ticau, S., Friedman, L.J., Ivica, N.A., Gelles, J., and Bell, S.P. (2015). Single-molecule studies of origin licensing reveal mechanisms ensuring bidirectional helicase loading. Cell *161*, 513-525.

Triolo, T., and Sternglanz, R. (1996). Role of interactions between the origin recognition complex and SIR1 in transcriptional silencing. Nature *381*, 251-253.

Tsukagoshi, H., Busch, W., and Benfey, P.N. (2010). Transcriptional regulation of ROS controls transition from proliferation to differentiation in the root. Cell *143*, 606-616.

Ubeda-Tomas, S., Federici, F., Casimiro, I., Beemster, G.T., Bhalerao, R., Swarup, R., Doerner, P., Haseloff, J., and Bennett, M.J. (2009). Gibberellin signaling in the endodermis controls Arabidopsis root meristem size. Curr Biol *19*, 1194-1199.

Valton, A.L., Hassan-Zadeh, V., Lema, I., Boggetto, N., Alberti, P., Saintome, C., Riou, J.F., and Prioleau, M.N. (2014). G4 motifs affect origin positioning and efficiency in two vertebrate replicators. EMBO J *33*, 732-746.

van den Berg, C., Willemsen, V., Hendriks, G., Weisbeek, P., and Scheres, B. (1997). Short-range control of cell differentiation in the Arabidopsis root meristem. Nature *390*, 287-289.

Vanstraelen, M., Baloban, M., Da Ines, O., Cultrone, A., Lammens, T., Boudolf, V., Brown, S.C., De Veylder, L., Mergaert, P., and Kondorosi, E. (2009). APC/C-CCS52A complexes control meristem maintenance in the Arabidopsis root. Proc Natl Acad Sci U S A *106*, 11806-11811.

Vas, A., Mok, W., and Leatherwood, J. (2001). Control of DNA rereplication via Cdc2 phosphorylation sites in the origin recognition complex. Mol Cell Biol *21*, 5767-5777.

Walter, J.C. (2000). Evidence for sequential action of cdc7 and cdk2 protein kinases during initiation of DNA replication in Xenopus egg extracts. J Biol Chem *275*, 39773-39778.

Wang, C., Liu, C., Roqueiro, D., Grimm, D., Schwab, R., Becker, C., Lanz, C., and Weigel, D. (2015). Genome-wide analysis of local chromatin packing in Arabidopsis thaliana. Genome Res *25*, 246-256.

Wang, H., Wang, J., Jiang, J., Chen, S., Guan, Z., Liao, Y., and Chen, F. (2014). Reference genes for normalizing transcription in diploid and tetraploid Arabidopsis. Sci Rep *4*, 6781.

Wang, J., Marowsky, N.C., and Fan, C. (2013). Divergent evolutionary and expression patterns between lineage specific new duplicate genes and their parental paralogs in Arabidopsis thaliana. PLoS One *8*, e72362.

Weimer, A.K., and al, e. (2016). The plant-specific CDKB1-CYCB1 complex mediates homologous recombination repair in Arabidopsis. EMBO J.

Weinberg, R.A. (1995). The retinoblastoma protein and cell cycle control. Cell *81*, 323-330.

Weingartner, M., Criqui, M.C., Meszaros, T., Binarova, P., Schmit, A.C., Helfer, A., Derevier, A., Erhardt, M., Bogre, L., and Genschik, P. (2004). Expression of a nondegradable cyclin B1 affects plant development and leads to endomitosis by inhibiting the formation of a phragmoplast. Plant Cell *16*, 643-657.

Weingartner, M., Pelayo, H.R., Binarova, P., Zwerger, K., Melikant, B., de la Torre, C., Heberle-Bors, E., and Bogre, L. (2003). A plant cyclin B2 is degraded early in mitosis and its ectopic expression shortens G2-phase and alleviates the DNA-damage checkpoint. J Cell Sci *116*, 487-498.

West, P.T., Li, Q., Ji, L., Eichten, S.R., Song, J., Vaughn, M.W., Schmitz, R.J., and Springer, N.M. (2014). Genomic distribution of H3K9me2 and DNA methylation in a maize genome. PLoS One *9*, e105267.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O*., et al.* (2007). A unified classification system for eukaryotic transposable elements. Nat Rev Genet *8*, 973-982.

Wierzbicki, A.T., Haag, J.R., and Pikaard, C.S. (2008). Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes. Cell *135*, 635-648.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science *290*, 2309-2312.

Wolff, C.F. (1774). Theoria generationis editio nova.

Wu, J.R., and Gilbert, D.M. (1996). A distinct G1 step required to specify the Chinese hamster DHFR replication origin. Science *271*, 1270-1272.

Xiao, J., Jin, R., and Wagner, D. (2017). Developmental transitions: integrating environmental cues with hormonal signaling in the chromatin landscape in plants. Genome Biol *18*, 88.

Xiao, J., and Wagner, D. (2015). Polycomb repression in the regulation of growth and development in Arabidopsis. Curr Opin Plant Biol *23*, 15-24.

Xu, H., Nelson, A.D., and Shippen, D.E. (2015). A transposable element within the Non-canonical telomerase RNA of Arabidopsis thaliana modulates telomerase in response to DNA damage [corrected]. PLoS Genet *11*, e1005281.

Yeeles, J.T., Deegan, T.D., Janska, A., Early, A., and Diffley, J.F. (2015). Regulated eukaryotic DNA replication origin firing with purified proteins. Nature *519*, 431-435.

You, Z., and Masai, H. (2008). Cdt1 forms a complex with the minichromosome maintenance protein (MCM) and activates its helicase activity. J Biol Chem *283*, 24469-24477.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952-1958.

Zegerman, P., and Diffley, J.F. (2007). Phosphorylation of Sld2 and Sld3 by cyclin-dependent kinases promotes DNA replication in budding yeast. Nature *445*, 281-285.

Zemach, A., Kim, M.Y., Hsieh, P.H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193-205.

Zhai, Y., Cheng, E., Wu, H., Li, N., Yung, P.Y., Gao, N., and Tye, B.K. (2017). Open-ringed structure of the Cdt1-Mcm2-7 complex as a precursor of the MCM double hexamer. Nat Struct Mol Biol *24*, 300-308.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E*., et al.* (2006). Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell *126*, 1189-1201.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W*., et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Zhao, X., Harashima, H., Dissmeyer, N., Pusch, S., Weimer, A.K., Bramsiepe, J., Bouyer, D., Rademacher, S., Nowack, M.K., Novak, B*., et al.* (2012). A general G1/S-phase cell-cycle control module in the flowering plant Arabidopsis thaliana. PLoS Genet *8*, e1002847.

Zhu, Y., Dong, A., Meyer, D., Pichon, O., Renou, J.P., Cao, K., and Shen, W.H. (2006). Arabidopsis NRP1 and NRP2 encode histone chaperones and are required for maintaining postembryonic root growth. Plant Cell *18*, 2879-2892.

Supplementary Material

Supplementary Table 1: Primers used for cloning.

Primer name	Sequence
pORC1a(-1324)::ORC1a F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGACCTCTAGTAGCATATGCC
pORC1a(-87)::ORC1a F	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGTGCAAACATTTCCCGCC
pORC1a::ORC1a R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAAGTAATTGGCCAACCATG
pORC1b(-793)::ORC1b F	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTGGGTTTCCAGAGAGTGG
pORC1b::ORC1b R	GGGGACCACTTTGTACAAGAAAGCTGGGTCCAAGTAATTGGCCAGCCA

Supplementary Table 2: Primers used for characterization of T-DNA insertion mutant lines.

Primer name	Sequence
ORC1a LP	TTTACGAGAATTTGCACCACC
ORC1a RP	CCTTCTGCGTCACCATCTTAG
ORC1a T-DNA	AACGTCCGCAATGTGTTATTAAGTTGTC
ORC1b LP	TCCTACATTCACCAAGATCGC
ORC1b RP	CGTAAAGGGAGATTCTTTGGG
ORC1b T-DNA	ATTTTGCCGATTTCGGAAC
FBL17-1 F	GAACTGCTTGATCTGAGTGGG
FBL17-1 R	CCAACTTCCTTCTCTTCCCTG
FBL17-1 T-DNA	CCCATTTGGACGTGAATGTAGA
SKP2a-1 LP	GCCTGAAGGATACAAGCACAG
SKP2a-1 RP	CCCAAGTTTGTAAAGCTGCAG
SKP2a-1 T-DNA	CCCATTTGGACGTGAATGTAG
ATXR5 LP	CCATTGGAACTTGGCTTTGTGTC
ATXR5 RP	AATAGGACCATCTGCTTCAACTGTG
ATXR5 T-DNA	TGGTTCACGTAGTGGGCCATCG
ATXR6 LP	AGCTTTGCTGGTTGTTTACCGGA
ATXR6 RP	CCATGTTGAGTAAATGTCGAAGAC
ATXR6 T-DNA	AACGTCCGCAATGTGTTATTAAGTTGTC

Supplementary Table 3: Primers used for qPCR analysis.

Primer name	Sequence	SNS	RNA	ChIP
AT1TE62820 1F	GTCGACAAGCCAAACTGGAT	x		
AT1TE62820 1R	CCGATTTTCGGTTCTCCATA			

Primer name	Sequence	SNS	RNA	ChIP
AT1TE62820 2F	TCCCGACAACAAAGTGAAGA	x		
AT1TE62820 2R	CAAATAAACGGGAGGGTGTG			
AT1TE62820 3F	ACTGCGGAAAACGCTACTCA	x	x	x
AT1TE62820 3R	ACATCGACCGGGAACATAGC			
AT1TE62820 4F	TCCGAATTGGTTCTCTCCAG	x		
AT1TE62820 4R	TGTCACCACTCGAGCTATGC			
AT1TE62820 5F	TTGCTTCGGAAACTTTCGCC	x	x	x
AT1TE62820 5R	ACCTCGTAGAGCGGGTGATA			
AT1TE62820 6F	TCTTCACCGACTACCGCATA	x		
AT1TE62820 6R	CGATCGGAACGAAATCCTAA			
AT1G51350 1F	AACACAATACCACAAACCAAAG	x		
AT1G51350 1R	AGTCAATGGAGTATAGATAGAG			
AT1G51350 2F	TTCCAATCTAAGCCAAAACTC	x		
AT1G51350 2R	ATCAGAATCGTCAGCATCAGC			
AT1G51350 3F	GTATTATCGCTCATGCTTGTG	x		
AT1G51350 3R	TGACAACTAAGCAAAGACAAG			
AT1G51350 4F	TCAATGGATCCAAATACTCGG	x		
AT1G51350 4R	TCAACAAGATTACGGAGGAGG			
AT1G51350 5F	CTTTACTTGTGCCACTTTTAGA	x		
AT1G51350 5R	ATTTTAATTTTATGTTTTGCCACG			
AT1G51350 6F	GTGGGTTTGAATTTCTGGTAG	x		
AT1G51350 6R	TGTACCCAATAAAAAGGAAATG			
AT4TE16725 1F	TGAAAAGCACTACTGCGCTAA	x		
AT4TE16725 1R	TCAACCGGACTGTTTGTTCA			
AT4TE16725 2F	CAGATCGGAAAAGGGAAGAG	x	x	x
AT4TE16725 2R	CCAAGGTAATTTCCCTCCTTC			
AT4TE16725 3F	CCCGGATTCCAGTTCACCTC	x	x	x
AT4TE16725 3R	TGGAGGCCATTAACGTGGAC			
AT4TE16725 4F	CACCGCGTGAACGAGATAGT	x		
AT4TE16725 4R	GCTACAGGACCATGTGAACCA			
AT4TE16725 5F	CAGTCAAAGGGGCAACCTAC	x		
AT4TE16725 5R	AAGTGTGCCTTGACCTTTGG			
AT4TE16725 6F	CTACAATGCCCTGCTCATCA	x		
AT4TE16725 6R	GCACCGGAGTCGTCAGTTAT			
AT2TE15565 1F	AAAGTGAAACCGGGTCAAAA	x		
AT2TE15565 1R	TCACAAAAATACAAGTGGGTTAAA			

Primer name	Sequence	SNS	RNA	ChIP
AT2TE15565 2F	TCCTCACACGCTCCACATTA	x	x	x
AT2TE15565 2R	GAAACCTGAGGCTGAGGAGA			
AT2TE15565 3F	ACCCGGACTCCTCTTAGGAC	x	x	x
AT2TE15565 3R	AAGGCACAGTGGTGGTGAAT			
AT2TE15565 4F	CTCGAACCAGTCATCTCATGG	x		
AT2TE15565 4R	GTGGTGGCAGTGGTCAAGA			
AT4TE03295 1F	TTGAAGAGTCCCCATTCCAC	x		
AT4TE03295 1R	CGCGCCAATTTACTGGTATT			
AT4TE03295 2F	TGAACTGGACCGGCTACTTT	x		
AT4TE03295 2R	GGTCCCCGTTAGGAAACAAG			
AT4TE03295 3F	GTAAACCACCTGCGACCATT	x	x	x
AT4TE03295 3R	TGAGCTGCGATACTGTCCAC			
AT4TE03295 4F	AGTCTAGCACGAGAGTGGCT	x	x	x
AT4TE03295 4R	TTCTTTGTTCCAGGCAGCGT			
AT4TE03295 5F	TTCTGCGGAGAGGAGGAGTC	x		
AT4TE03295 5R	GGTATTGGGATCCAAGGGCT			
AT4TE03295 6F	GTGGCTACATCCCTGACGAT	x		
AT4TE03295 6R	ACGGTGGAAAACTCCGGTAT			
AT2TE13970 F	TTCGGCTTGATTTGAGCCAC		x	x
AT2TE13970 R	GTTAGCTGGTGGGAGGACAG			
AT2TE16335 F	TGTGAAAGCCCATTACACTTTACT		x	x
AT2TE16335 R	ACGACGAATAAGGTTTGGGAGA			
AT4TE16735 F	TCGTCAAAATCTTTGGTGCTTGT		x	x
AT4TE16735 R	GCAATTAACGATCACCTAATCATGG			
AT4TE17050 F	ACACCAAATCTGCCACTCCA		x	x
AT4TE17050 R	TGTTTTTGGTGAATGATTGGATAGT			
GAPC2 F	TCCAACGCTAGTTGCACCAC		x	
GAPC2 R	TGGACAGTGGTCATGAGTCC			
CYCA3;1 F	GTACTCGGAGAGCTTCCAAACT		x	
CYCA3;1 R	CGAGGTAAAGAGTGTCTGAGAG			
CYCB1;4 F	GATCAATCATCGTCCTCGTACACG		x	
CYCB1;4 R	TGCCATGTAATCTCGTGGCCTCC			
BRCA1 F	CATGTGCCTTTTGTCAGTGTTC		x	
BRCA1 R	TGGAGCCCATTCAGCACAGTTT			
PARP1 F	CTCCTGAAGCGCCTGTAACT		x	
PARP1 R	CATGTCTCCCAAAGCAACCT			

Primer name	Sequence	SNS	RNA	ChIP
RAD51 F	GCTAGTTCCTCTGGGGTTCA		x	
RAD51 R	GGAACCAGTTTCAATACCTCCT			
AT4TE37370 F	CTTCAAAAGAATCTCTTACCC		x	
AT4TE37370 R	ATTTATAGAAAGACATTAAATTTGG			

Supplementary Table 4: Primers used for evaluation of mRNA in the T-DNA insertion mutant lines by PCR

Primer name	Sequence
ORC1a 1	ATGGCTTCTTCTCTGAGTTCCA
ORC1a 2	AACGTCCGCAATGTGTTATTAAGTTGTC
ORC1a 3	CAGGTATCATATACCAACGAGCTC
ORC1b 4	GCTGATGGAGATTCTG
ORC1b 5	CTGAATTGGCTGATGGAGATTCTG
ORC1b 6	ATTTTGCCGATTTCGGAAC
ORC1b 7	GAATAAATATTCTCTAGAGAA

Appendix

Publications derived from the doctoral thesis:

Vergara, Z., Sequeira-Mendes, J., Morata, J., Peiro, R., Henaff, E., Costas, C., Casacuberta, J.M. & Gutierrez, C. (2017) Retrotransposons are specified as DNA replication origins in the gene-poor regions of Arabidopsis heterochromatin. ***Nucleic Acids Res*** doi: https://doi.org/10.1093/nar/gkx524

Desvoyes, B., Vergara, Z., Sequeira-Mendes, J., Madeira, S., Gutierrez, C. (2017) A rapid and efficient ChIP protocol to profile chromatin binding proteins and epigenetic modifications in bulk Arabidopsis tissue. ***Methods Mol. Biol.*** (in press).

Desvoyes, B., Sequeira-Mendes, J., Vergara, Z., Madeira, S., Gutierrez, C. (2017) Sequential ChIP protocol for profiling bivalent epigenetic modifications (ReChIP). ***Methods Mol. Biol.*** (in press).

Vergara, Z. & Gutierrez, C. (2017) Emerging roles of chromatin in the maintenance of genome organization and function in plants. ***Genome Biol.*** 18, 96.

Gutierrez, C., Desvoyes, B., Vergara, Z., Otero, S. & Sequeira-Mendes, J. (2016) Links of genome replication, transcriptional silencing and chromatin dynamics. ***Curr Opin Plant Biol*** 34**,** 92–99.

Desvoyes, B., Fernandez-Marcos, M., Sequeira-Mendes, J., Otero, S., Vergara, Z. & Gutierrez, C. (2014) Looking at plant cell cycle from the chromatin window. ***Front Plant Sci*** 5**,** 369.